

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:33:43 ; Search time 45 Seconds  
(without alignments)  
2455.024 Million cell updates/sec

Title: US-09-854-122-16  
Perfect score: 2059  
Sequence: 1 TRGITGIAVACLEPLIMAGIL.....CLLSYSFVINPILLKKKK 391

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A Geneseq\_29Jan04:\*  
1: Geneseqp1980s:\*  
2: Geneseqp1990s:\*  
3: Geneseqp2000s:\*  
4: Geneseqp2001s:\*  
5: Geneseqp2002s:\*  
6: Geneseqp2003as:\*  
7: Geneseqp2003bs:\*  
8: Geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1996	96.9	397	5 AAU12056	Aau12056 Protein c
2	1750	85.0	331	5 AAU12057	Aau12057 Zostera m
3	572.5	27.8	331	3 AAG21657	Aag21657 Arabidops
4	572.5	27.8	341	3 AAG21656	Aag21656 Arabidops
5	567	27.5	346	5 ABB91056	Abb91056 Arabidops
6	567	27.5	346	6 ABB91056	Abb91056 Arabidops
7	562.5	27.3	338	6 ABB91056	Abb91056 Arabidops
8	561.5	27.3	338	3 AAG16972	Aag16972 Arabidops
9	558.5	27.1	350	3 AAG17409	Aag17409 Arabidops
10	558.5	27.1	350	5 ABB91589	Abb91589 Arabidops
11	558.5	27.1	350	6 ABB91589	Abb91589 Arabidops
12	555.5	27.0	359	3 AAG46836	Aag46836 Arabidops
13	555.5	27.0	359	3 AAG44205	Aag44205 Arabidops
14	551.5	26.8	359	4 AAB49722	Aab49722 Hydroxyja
15	551.5	26.8	354	3 AAG44206	Aag44206 Arabidops
16	551.5	26.8	354	3 AAG46837	Aag46837 Arabidops
17	551.5	26.8	354	5 ABB93382	Abb93382 Arabidops
18	550.5	26.7	327	3 AAG16879	Aag16879 Arabidops
19	550.5	26.7	337	3 AAG16878	Aag16878 Arabidops
20	550.5	26.7	349	3 AAG46838	Aag46838 Arabidops
21	550.5	26.7	349	3 AAG44207	Aag44207 Arabidops
22	548.5	26.6	351	3 AAG17285	Aag17285 Arabidops
23	545	26.5	318	3 AAG17287	Aag17287 Arabidops
24	545	26.5	331	3 AAG17286	Aag17286 Arabidops
25	545	26.5	331	5 ABB91710	Abb91710 Arabidops

26	544.5	26.4	326	5 ABB91155	Abb91155 Arabidops
27	537.5	26.1	326	3 AAG29670	Aag29670 Arabidops
28	537.5	26.1	326	5 ABB91711	Abb91711 Arabidops
29	537.5	26.1	326	6 ABB91192	Abb91192 Arabidops
30	536.5	26.1	351	5 ABB90981	Abb90981 Arabidops
31	525.5	25.5	333	5 ABB91753	Abb91753 Arabidops
32	510.5	24.8	340	5 ABB93381	Abb93381 Arabidops
33	510.5	24.8	347	4 AAB49723	Aab49723 Hydroxyja
34	486.5	23.6	314	5 ABB93110	Abb93110 Arabidops
35	483	23.5	688	7 ADD25205	Add25205 Fertilility
36	407	19.8	273	5 ABB91902	Abb91902 Arabidops
37	393.5	19.1	258	6 AAE35941	Aae35941 Lolium pe
38	372	18.1	247	6 AAE35942	Aae35942 Lolium pe
39	343	16.7	201	3 AAG16880	Aag16880 Arabidops
40	331	16.1	190	3 AAG29671	Aag29671 Arabidops
41	321	15.6	179	3 AAG16973	Aag16973 Arabidops
42	317	15.4	216	3 AAG04466	Aag04466 Arabidops
43	314	15.3	186	3 AAG21658	Aag21658 Arabidops
44	301	14.6	179	3 AAG17410	Aag17410 Arabidops
45	292.5	14.2	226	6 AAE35947	Aae35947 Lolium pe

ALIGNMENTS

RESULT 1  
AAU12056  
ID AAU12056 standard; protein; 397 AA.

AC AAU12056;

DT 09-APR-2002 (first entry)

DE Protein containing sulfotransferase encoded by Z. marina cDNA clone.  
KW Plant; transgenic; marine eelgrass; zosteric acid biosynthesis;  
KW saline-resistance; anoxia-resistance; anti-fouling genetic trait;  
KW marine vascular plant; sulphated phenolic compound; Zostera marina;  
KW sulfotransferase; S1; enzyme.

OS Zostera marina.

XX Key Location/Qualifiers

FT Misc-difference 12 /label= Xaa /note= "Xaa= Stop codon, encoded by TAG"

FT Protein 17. .347 /note= "Sulfotransferase"

FT Misc-difference 348 /label= Xaa /note= "Xaa= Stop codon, encoded by TAA"

FT Misc-difference 354 /label= Xaa /note= "Xaa= Stop codon, encoded by TGA"

FT Misc-difference 360 /label= Xaa /note= "Xaa= Stop codon, encoded by TGA"

FT Misc-difference 363 /label= Xaa /note= "Xaa= Stop codon, encoded by TGA"

FT Misc-difference 374 /label= Xaa /note= "Xaa= Stop codon, encoded by TAA"

XX WO200185971-A2.

XX 15-NOV-2001.

XX 10-MAY-2001; 2001WO-US015412.

XX 10-MAY-2000; 2000US-0202529P.

XX (PHYC-) PHYCOGEN INC.

PA

XX PI Alberte RS, Smith RD;  
 XX WPI; 2002-121947/16.  
 XX N-PSDB; AAS20863.  
 XX PT New transgenic plants comprising a zosteric acid biosynthetic gene, a  
 PT saline resistance gene or a hypoxia resistance gene derived from *Zostera*  
 PT marina, useful for producing plants with antifouling traits.  
 XX XX  
 PS Example; Fig 4; 117pp; English.  
 XX XX  
 CC The present invention relates to a new transgenic plant comprising a  
 CC heterologous gene derived from the marine eelgrass *Zostera marina* or at  
 CC least one heterologous nucleotide sequence encoding a zosteric acid  
 CC biosynthetic function, a saline-resistance function, or a anoxia-  
 CC resistance function. The invention describes the method of producing a  
 CC transgenic plant possessing an anti-fouling genetic trait by providing a  
 CC cDNA population derived from a marine vascular plant, isolating from the  
 CC cDNA population a nucleic acid species which hybridises to a nucleic acid  
 CC that encodes a sulfotransferase (ST), an alcohol dehydrogenase (ADH),  
 CC phenylalanine ammonia lyase (PAL) or a cinnamate-4-hydroxylase (CH), and  
 CC transforming a target host plant with the isolated nucleic acid. The  
 CC plant is useful in the genetic engineering of plant species having  
 CC desirable genetic traits such as antifouling traits, salt and anoxia  
 CC resistance, and pathogen defence strategy. The expression of such  
 CC biosynthetic enzymes are sufficient to support the production of zosteric  
 CC acid and other sulphated phenolic compounds in a target plant. The  
 CC present sequence represents a protein containing sulfotransferase which  
 CC is encoded by a *Z. marina* cDNA clone  
 XX XX  
 SQ Sequence 397 AA;  
 Query Match 96.9%; Score 1996; DB 5; Length 397;  
 Best Local Similarity 98.5%; Pred. No. 3.4e-166;  
 Matches 391; Conservative 0; Mismatches 0; Indels 6; Gaps 6;  
 QY 1 TRGITGAVAC-LPLIMAGILALEKCFGSKNEQKEEDSKMYRYREIVSSLPNSNDYWD 59  
 DB 1 TRGITGAVACXPLIMAGILALEKCFGSKNEQKEEDSKMYRYREIVSSLPNSNDYWD 60  
 QY 60 TMRLYKFGWQVYVPGIMAFENFKARETDIILTLPKAGTTWKALTFAILLTRDVNHP 119  
 DB 61 TMRLYKFGWQVYVPGIMAFENFKARETDIILTLPKAGTTWKALTFAILLTRDVNHP 120  
 QY 120 SSPTHPLFFNPHSCVQNLLEYLYMGRENTMPDLMLNESPRLPFAGHIPYSLLPASVLKSG 179  
 DB 121 SSPTHPLFFNPHSCVQNLLEYLYMGRENTMPDLMLNESPRLPFAGHIPYSLLPASVLKSG 180  
 QY 180 TKIINISRNKSTFVSFWKFGNLIINDKLLDLEKSVDFIFASGISFCGPEWNFQAEFTAA 239  
 DB 181 TKIINISRNKSTFVSFWKFGNLIINDKLLDLEKSVDFIFASGISFCGPEWNFQAEFTAA 240  
 QY 240 STNSNLLLSYEEMLEKFPVENVKLAELFMGCGGFTDDEEKQGIIVDEIVKLCSFDNLKNQV 299  
 DB 241 STNSNLLLSYEEMLEKFPVENVKLAELFMGCGGFTDDEEKQGIIVDEIVKLCSFDNLKNQV 300  
 QY 300 NKGSSVNSKIDNKHFRKGEVDWANYLTSEMVKLETAGINSE-KHLIS-NKNTL- 356  
 DB 301 NKGSSVNSKIDNKHFRKGEVDWANYLTSEMVKLETAGINSEKXHLISXKNNTX 360  
 QY 357 NF-NNLNNYCKSN-ISLCLLSYSFVINNFILKXKKK 391  
 DB 361 NFXNLNNYCKSNISLCLLSYSFVINNFILKXKKK 397  
 RESULT 2  
 AAU12057  
 ID AAU12057 standard; protein; 331 AA.  
 XX  
 AC AAU12057;  
 XX  
 DT 09-APR-2002 (first entry)

XX Zostera marina sulfotransferase protein.  
 XX DE  
 XX Plant; transgenic; marine eelgrass; zosteric acid biosynthesis;  
 XX KW saline-resistance; anoxia-resistance; anti-fouling genetic trait;  
 XX KW marine vascular plant; sulphated phenolic compound; *Zostera marina*;  
 XX KW sulfotransferase; ST; enzyme.  
 XX OS  
 XX Zostera marina.  
 XX PN WO200185971-A2.  
 XX XX  
 XX 15-NOV-2001.  
 XX PD  
 XX 10-MAY-2001; 2001WO-US015412.  
 XX PF  
 XX 10-MAY-2000; 2000US-020529P.  
 XX PR  
 XX (PHYC-) PHYCOGEN INC.  
 XX PA  
 XX Alberte RS, Smith RD;  
 XX PI  
 XX WPI; 2002-121947/16.  
 XX DR  
 XX New transgenic plants comprising a zosteric acid biosynthetic gene, a  
 XX saline resistance gene or a hypoxia resistance gene derived from *Zostera*  
 XX marina, useful for producing plants with antifouling traits.  
 XX PT  
 XX Example; Fig 5; 117pp; English.  
 XX PS  
 XX XX  
 CC The present invention relates to a new transgenic plant comprising a  
 CC heterologous gene derived from the marine eelgrass *Zostera marina* or at  
 CC least one heterologous nucleotide sequence encoding a zosteric acid  
 CC biosynthetic function, a saline-resistance function, or a anoxia-  
 CC resistance function. The invention describes the method of producing a  
 CC transgenic plant possessing an anti-fouling genetic trait by providing a  
 CC cDNA population derived from a marine vascular plant, isolating from the  
 CC cDNA population a nucleic acid species which hybridises to a nucleic acid  
 CC that encodes a sulfotransferase (ST), an alcohol dehydrogenase (ADH),  
 CC phenylalanine ammonia lyase (PAL) or a cinnamate-4-hydroxylase (CH), and  
 CC transforming a target host plant with the isolated nucleic acid. The  
 CC plant is useful in the genetic engineering of plant species having  
 CC desirable genetic traits such as antifouling traits, salt and anoxia  
 CC resistance, and pathogen defence strategy. The expression of such  
 CC biosynthetic enzymes are sufficient to support the production of zosteric  
 CC acid and other sulphated phenolic compounds in a target plant. The  
 CC present sequence represents *Z. marina* sulfotransferase protein  
 XX XX  
 SQ Sequence 331 AA;  
 Query Match 85.0%; Score 1750; DB 5; Length 331;  
 Best Local Similarity 100.0%; Pred. No. 9.6e-145;  
 Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 16 MAGILALEKCFGSKNEQKEEDSKMYRYREIVSSLPNSNDYWDITMRLYKFGWQVYLP 75  
 DB 1 MAGILALEKCFGSKNEQKEEDSKMYRYREIVSSLPNSNDYWDITMRLYKFGWQVYLP 60  
 QY 76 GIMAFEDNFKARETDIILTLPKAGTTWKALTFAILLTRDVNHPSSPTHLFFNPHSCV 135  
 DB 61 GIMAFEDNFKARETDIILTLPKAGTTWKALTFAILLTRDVNHPSSPTHLFFNPHSCV 120  
 QY 136 QNLEYLYMGRENTMPDLMLNESPRLPFAGHIPYSLLPASVLKSGTKIINISRNKSTFVS 195  
 DB 121 QNLEYLYMGRENTMPDLMLNESPRLPFAGHIPYSLLPASVLKSGTKIINISRNKSTFVS 180  
 QY 196 FWKFGNLIINDKLLDLEKSVDFIFASGISFCGPEWNFQAEFTAASTNSNLLLSYEEMLE 255  
 DB 181 FWKFGNLIINDKLLDLEKSVDFIFASGISFCGPEWNFQAEFTAASTNSNLLLSYEEMLE 240  
 QY 256 KPVENVKLAELFMGCGGFTDDEEKQGIIVDEIVKLCSFDNLKNQVNSKIDNKH 315  
 DB 241 KPVENVKLAELFMGCGGFTDDEEKQGIIVDEIVKLCSFDNLKNQVNSKIDNKH 300

Qy 316 FRKEVRDWNLYTSEMIMKLETAGKINESE 346  
Db 301 FRKEVRDWNLYTSEMIMKLETAGKINESE 331

## RESULT 3

AAG21657  
ID AAG21657 standard; protein; 331 AA.

XX AC AAG21657;

XX DT 17-OCT-2000 (first entry)

XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 24291.

XX KW Protein identification; signal transduction pathway; metabolic pathway;  
XX KW hybridisation assay; genetic mapping; gene expression control; promoter;  
XX KW termination sequence.

OS Arabidopsis thaliana.

XX FN EP1033405-A2.

XX PD 06-SEP-2000.

XX PF 25-FEB-2000; 2000EP-00301439.

XX PR 25-FEB-1999; 99US-0121825P.

PR 05-MAR-1999; 99US-0123180P.

PR 09-MAR-1999; 99US-0123548P.

PR 23-MAR-1999; 99US-0125788P.

PR 25-MAR-1999; 99US-0126264P.

PR 29-MAR-1999; 99US-0126785P.

PR 01-APR-1999; 99US-0127462P.

PR 06-APR-1999; 99US-0128234P.

PR 08-APR-1999; 99US-0128714P.

PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-0130077P.

PR 21-APR-1999; 99US-0130449P.

PR 23-APR-1999; 99US-0130510P.

PR 28-APR-1999; 99US-0130891P.

PR 30-APR-1999; 99US-0131449P.

PR 30-APR-1999; 99US-0132048P.

PR 04-MAY-1999; 99US-0132407P.

PR 05-MAY-1999; 99US-0132485P.

PR 06-MAY-1999; 99US-0132486P.

PR 07-MAY-1999; 99US-0132487P.

PR 11-MAY-1999; 99US-0132863P.

PR 14-MAY-1999; 99US-0134256P.

PR 14-MAY-1999; 99US-0134218P.

PR 14-MAY-1999; 99US-0134219P.

PR 14-MAY-1999; 99US-0134221P.

PR 18-MAY-1999; 99US-0134370P.

PR 19-MAY-1999; 99US-0134768P.

PR 19-MAY-1999; 99US-0134941P.

PR 17-JUN-1999; 99US-0139492P.

PR 18-JUN-1999; 99US-0139454P.

PR 18-JUN-1999; 99US-0139455P.

PR 18-JUN-1999; 99US-0139456P.

PR 18-JUN-1999; 99US-0139457P.

PR 18-JUN-1999; 99US-0139458P.

PR 18-JUN-1999; 99US-0139459P.

PR 18-JUN-1999; 99US-0139460P.

PR 18-JUN-1999; 99US-0139461P.

PR 18-JUN-1999; 99US-0139462P.

PR 18-JUN-1999; 99US-0139463P.

PR 18-JUN-1999; 99US-0139750P.

PR 18-JUN-1999; 99US-0139763P.

PR 21-JUN-1999; 99US-0139817P.

PR 22-JUN-1999; 99US-0139899P.

PR 23-JUN-1999; 99US-0140353P.

PR 23-JUN-1999; 99US-0140354P.

PR 24-JUN-1999; 99US-0140695P.

PR 28-JUN-1999; 99US-0140823P.

PR 29-JUN-1999; 99US-0140991P.

PR 30-JUN-1999; 99US-0141287P.

PR 01-JUL-1999; 99US-0141842P.

PR 01-JUL-1999; 99US-0142154P.

PR 02-JUL-1999; 99US-0142055P.

PR 06-JUL-1999; 99US-0142390P.

PR 08-JUL-1999; 99US-0142803P.

PR 09-JUL-1999; 99US-0142920P.

PR 12-JUL-1999; 99US-0142977P.

PR 13-JUL-1999; 99US-0143542P.

PR 14-JUL-1999; 99US-0143624P.

PR 15-JUL-1999; 99US-0144005P.

PR 16-JUL-1999; 99US-0144085P.

PR 16-JUL-1999; 99US-0144086P.

PR 17-JUL-1999; 99US-0144325P.

PR 19-JUL-1999; 99US-0144331P.

PR 19-JUL-1999; 99US-0144332P.

PR 19-JUL-1999; 99US-0144333P.

PR 19-JUL-1999; 99US-0144334P.

PR 19-JUL-1999; 99US-0144335P.

PR 20-JUL-1999; 99US-0144352P.

PR 20-JUL-1999; 99US-0144632P.

PR 20-JUL-1999; 99US-0144884P.

PR 21-JUL-1999; 99US-0144814P.

PR 21-JUL-1999; 99US-0145086P.

PR 21-JUL-1999; 99US-0145088P.

PR 22-JUL-1999; 99US-0145085P.

PR 22-JUL-1999; 99US-0145087P.

PR 22-JUL-1999; 99US-0145089P.

PR 23-JUL-1999; 99US-0145192P.

PR 23-JUL-1999; 99US-0145145P.

PR 23-JUL-1999; 99US-0145218P.

PR 23-JUL-1999; 99US-0145224P.

PR 26-JUL-1999; 99US-0145276P.

PR 27-JUL-1999; 99US-0145913P.

PR 27-JUL-1999; 99US-0145918P.

PR 27-JUL-1999; 99US-0145919P.

PR 28-JUL-1999; 99US-0145951P.

PR 02-AUG-1999; 99US-0146386P.

PR 02-AUG-1999; 99US-0146388P.

PR 02-AUG-1999; 99US-0146389P.

PR 02-AUG-1999; 99US-0147038P.

PR 03-AUG-1999; 99US-0147204P.

PR 04-AUG-1999; 99US-0147302P.

PR 04-AUG-1999; 99US-0147302P.

PR 05-AUG-1999; 99US-0147302P.

PR 05-AUG-1999; 99US-0147260P.

PR 06-AUG-1999; 99US-0147303P.

PR 06-AUG-1999; 99US-0147166P.

PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.

PR 10-AUG-1999; 99US-0148171P.

PR 11-AUG-1999; 99US-0148319P.

PR 12-AUG-1999; 99US-0148341P.

PR 13-AUG-1999; 99US-0148565P.

PR	13-AUG-1999;	99US-0148684P.	Db	72	DIIVASFPKSGTTWKALTFALAORSKH--TSDNHPLLTENPHELVPYLELDLVL--KSS	127
PR	16-AUG-1999;	99US-0149368P.	QY	149	MPDLML-NESPRLPAGHIPYSLLPASVLKSGTKIINISRNKSTFTVFWKFGNLPDK	207
PR	17-AUG-1999;	99US-0149175P.	Db	128	KPDLTKLPSPPRLPSTFMSFDALKVPLKSPCKIVYVCRNVNDVLISLWCFENSMGSEN	187
PR	20-AUG-1999;	99US-0149722P.	QY	208	ILDLEKSVDIIFASGISFCGPEW-NFOAFTNAASTNSNLLLSYBEMLEKPVENVKLA	266
PR	20-AUG-1999;	99US-0149723P.	Db	188	NLSLEALFESLCGYNLCPLWENVLGYWRGSLDPKHVLFRLRYEELKTEPRVQIKLA	247
PR	23-AUG-1999;	99US-0149902P.	QY	327	YLTSEMICKLE	337
PR	23-AUG-1999;	99US-0149920P.	Db	307	YMTPEMENKID	317
PR	25-AUG-1999;	99US-0150566P.				
PR	26-AUG-1999;	99US-0150884P.				
PR	27-AUG-1999;	99US-0151065P.				
PR	27-AUG-1999;	99US-0151066P.				
PR	27-AUG-1999;	99US-0151080P.				
PR	30-AUG-1999;	99US-0151303P.				
PR	31-AUG-1999;	99US-0151438P.				
PR	01-SEP-1999;	99US-0151930P.				
PR	07-SEP-1999;	99US-0152363P.				
PR	10-SEP-1999;	99US-0153070P.				
PR	13-SEP-1999;	99US-0153758P.				
PR	15-SEP-1999;	99US-0154018P.				
PR	16-SEP-1999;	99US-0154039P.				
PR	20-SEP-1999;	99US-0154779P.				
PR	22-SEP-1999;	99US-0155132P.				
PR	23-SEP-1999;	99US-0155486P.				
PR	24-SEP-1999;	99US-0155659P.				
PR	28-SEP-1999;	99US-0156458P.				
PR	29-SEP-1999;	99US-0156596P.				
PR	04-OCT-1999;	99US-0157117P.				
PR	05-OCT-1999;	99US-0157533P.				
PR	06-OCT-1999;	99US-0157865P.				
PR	08-OCT-1999;	99US-0158029P.				
PR	08-OCT-1999;	99US-0158232P.				
PR	12-OCT-1999;	99US-0158369P.				
PR	13-OCT-1999;	99US-0159293P.				
PR	13-OCT-1999;	99US-0159294P.				
PR	13-OCT-1999;	99US-0159295P.				
PR	14-OCT-1999;	99US-0159329P.				
PR	14-OCT-1999;	99US-0159330P.				
PR	14-OCT-1999;	99US-0159331P.				
PR	14-OCT-1999;	99US-0159637P.				
PR	14-OCT-1999;	99US-0159638P.				
PR	18-OCT-1999;	99US-0159584P.				
PR	21-OCT-1999;	99US-0160741P.				
PR	21-OCT-1999;	99US-0160767P.				
PR	21-OCT-1999;	99US-0160768P.				
PR	21-OCT-1999;	99US-0160770P.				
PR	21-OCT-1999;	99US-0160814P.				
PR	21-OCT-1999;	99US-0160815P.				
PR	22-OCT-1999;	99US-0160980P.				
PR	22-OCT-1999;	99US-0160981P.				
PR	22-OCT-1999;	99US-0160989P.				
PR	22-OCT-1999;	99US-0161404P.				
PR	23-OCT-1999;	99US-0161405P.				
PR	23-OCT-1999;	99US-0161406P.				
PR	26-OCT-1999;	99US-0161359P.				
PR	26-OCT-1999;	99US-0161360P.				
PR	26-OCT-1999;	99US-0161361P.				
PR	28-OCT-1999;	99US-0161920P.				
PR	28-OCT-1999;	99US-0161932P.				
PR	28-OCT-1999;	99US-0161933P.				
PR	29-OCT-1999;	99US-0162142P.				
Query Match 27.8%; Score 572.5; DB 3; Length 331;						
Best Local Similarity 39.5%; Pred. No. 1.7e-41;						
Matches 123; Conservative 61; Mismatches 118; Indels 9; Gaps 7;						
QY	31	EOEKEESKMKYKREIVSSLPN-DYWGDTMLRYKGFQMGWLVPQIMAFEDNFKARET	89			
Db	12	EEEEEEENQSEETKSLISSLPSSIDCGTKLYKQCWDKLLQNLINFNKQFQET	71			
QY	90	DIILTLTPKAGTTWKALTFALTRDYNHPSPSPHLLFFNPHSPCVQNLLEY-LYMGRENT	148			



PR 25-MAY-1999; 99US-0136021P.  
 PR 27-MAY-1999; 99US-0136392P.  
 PR 28-MAY-1999; 99US-0136782P.  
 PR 01-JUN-1999; 99US-0137222P.  
 PR 03-JUN-1999; 99US-0137528P.  
 PR 04-JUN-1999; 99US-0137502P.  
 PR 07-JUN-1999; 99US-0137724P.  
 PR 08-JUN-1999; 99US-0138094P.  
 PR 10-JUN-1999; 99US-0138540P.  
 PR 14-JUN-1999; 99US-0138847P.  
 PR 14-JUN-1999; 99US-0139119P.  
 PR 16-JUN-1999; 99US-0139452P.  
 PR 16-JUN-1999; 99US-0139453P.  
 PR 17-JUN-1999; 99US-0139492P.  
 PR 18-JUN-1999; 99US-0139454P.  
 PR 18-JUN-1999; 99US-0139455P.  
 PR 18-JUN-1999; 99US-0139456P.  
 PR 18-JUN-1999; 99US-0139457P.  
 PR 18-JUN-1999; 99US-0139476P.  
 PR 18-JUN-1999; 99US-0139458P.  
 PR 18-JUN-1999; 99US-0139459P.  
 PR 18-JUN-1999; 99US-0139460P.  
 PR 18-JUN-1999; 99US-0139461P.  
 PR 18-JUN-1999; 99US-0139462P.  
 PR 18-JUN-1999; 99US-0139463P.  
 PR 18-JUN-1999; 99US-0139750P.  
 PR 18-JUN-1999; 99US-0139763P.  
 PR 21-JUN-1999; 99US-0139817P.  
 PR 22-JUN-1999; 99US-0139899P.  
 PR 23-JUN-1999; 99US-0140353P.  
 PR 23-JUN-1999; 99US-0140354P.  
 PR 24-JUN-1999; 99US-0140695P.  
 PR 24-JUN-1999; 99US-0140823P.  
 PR 29-JUN-1999; 99US-0140931P.  
 PR 30-JUN-1999; 99US-0141287P.  
 PR 01-JUL-1999; 99US-0141842P.  
 PR 01-JUL-1999; 99US-0142154P.  
 PR 02-JUL-1999; 99US-0142055P.  
 PR 06-JUL-1999; 99US-0142390P.  
 PR 08-JUL-1999; 99US-0142803P.  
 PR 09-JUL-1999; 99US-0142920P.  
 PR 12-JUL-1999; 99US-0142977P.  
 PR 13-JUL-1999; 99US-0143542P.  
 PR 14-JUL-1999; 99US-0143624P.  
 PR 15-JUL-1999; 99US-0144005P.  
 PR 16-JUL-1999; 99US-0144085P.  
 PR 16-JUL-1999; 99US-0144086P.  
 PR 19-JUL-1999; 99US-0144325P.  
 PR 19-JUL-1999; 99US-0144331P.  
 PR 19-JUL-1999; 99US-0144332P.  
 PR 19-JUL-1999; 99US-0144333P.  
 PR 19-JUL-1999; 99US-0144334P.  
 PR 19-JUL-1999; 99US-0144335P.  
 PR 20-JUL-1999; 99US-0144352P.  
 PR 20-JUL-1999; 99US-0144632P.  
 PR 20-JUL-1999; 99US-0144884P.  
 PR 21-JUL-1999; 99US-0144814P.  
 PR 21-JUL-1999; 99US-0145086P.  
 PR 21-JUL-1999; 99US-0145088P.  
 PR 22-JUL-1999; 99US-0145085P.  
 PR 22-JUL-1999; 99US-0145087P.  
 PR 22-JUL-1999; 99US-0145089P.  
 PR 22-JUL-1999; 99US-0145192P.  
 PR 23-JUL-1999; 99US-0145145P.  
 PR 23-JUL-1999; 99US-0145218P.  
 PR 23-JUL-1999; 99US-0145224P.  
 PR 26-JUL-1999; 99US-0145276P.  
 PR 27-JUL-1999; 99US-0145911P.  
 PR 27-JUL-1999; 99US-0145918P.  
 PR 28-JUL-1999; 99US-0145951P.  
 PR 02-AUG-1999; 99US-0146386P.  
 PR 02-AUG-1999; 99US-0146388P.  
 PR 02-AUG-1999; 99US-0146389P.

PR 03-AUG-1999; 99US-0147038P.  
 PR 04-AUG-1999; 99US-0147204P.  
 PR 04-AUG-1999; 99US-0147302P.  
 PR 05-AUG-1999; 99US-0147192P.  
 PR 05-AUG-1999; 99US-0147260P.  
 PR 06-AUG-1999; 99US-0147303P.  
 PR 06-AUG-1999; 99US-0147416P.  
 PR 09-AUG-1999; 99US-0147493P.  
 PR 09-AUG-1999; 99US-0147935P.  
 PR 10-AUG-1999; 99US-0148171P.  
 PR 11-AUG-1999; 99US-0148319P.  
 PR 12-AUG-1999; 99US-0148341P.  
 PR 13-AUG-1999; 99US-0148565P.  
 PR 13-AUG-1999; 99US-0148684P.  
 PR 16-AUG-1999; 99US-0149368P.  
 PR 17-AUG-1999; 99US-0149175P.  
 PR 18-AUG-1999; 99US-0149426P.  
 PR 20-AUG-1999; 99US-0149722P.  
 PR 20-AUG-1999; 99US-0149723P.  
 PR 20-AUG-1999; 99US-0149923P.  
 PR 23-AUG-1999; 99US-0149902P.  
 PR 23-AUG-1999; 99US-0149930P.  
 PR 25-AUG-1999; 99US-0150566P.  
 PR 26-AUG-1999; 99US-0150884P.  
 PR 27-AUG-1999; 99US-0151065P.  
 PR 27-AUG-1999; 99US-0151066P.  
 PR 27-AUG-1999; 99US-0151080P.  
 PR 30-AUG-1999; 99US-0151303P.  
 PR 31-AUG-1999; 99US-0151438P.  
 PR 01-SEP-1999; 99US-0151930P.  
 PR 07-SEP-1999; 99US-0152363P.  
 PR 10-SEP-1999; 99US-0153070P.  
 PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154039P.  
 PR 20-SEP-1999; 99US-0154779P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 23-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155659P.  
 PR 28-SEP-1999; 99US-0156458P.  
 PR 28-SEP-1999; 99US-0156596P.  
 PR 04-OCT-1999; 99US-0157117P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157865P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 18-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160980P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 28-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161320P.  
 PR 28-OCT-1999; 99US-0161992P.



CC AtST5a and AtST5b. AtST5a and AtST5b are constitutively expressed in all  
 CC parts of the plants, while AtST5c does not seem to be expressed in the  
 CC early stage of development. These enzymes are involved in glucosinolate  
 CC synthesis in plants. The desulfoglucosinolate sulfotransferase  
 CC polynucleotides are useful for modulating glucosinolate biosynthesis in  
 CC transgenic plants  
 XX Sequence 346 AA;

Query Match 27.5%; Score 567; DB 6; Length 346;  
 Best Local Similarity 39.6%; Pred. NO. 5.5e-41;  
 Matches 137; Conservative 56; Mismatches 123; Indels 30; Gaps 11;  
 QY 23 EKCFGKNEQKEBDSKMYKRYREIVSSLPNDY--GDTWRLYKGFQWQGYLVPGIMAF 80  
 DB 17 ELASSPSEFEKQ-----KHVQELIATLPHKDGWRPKDPVEYGGHWLQPLLEGLLHA 71  
 QY 81 EDNFKARTDILITLPAKATTTWKALTFAILTR---DVNHPSPSTHPLFFNPHSCVQN 137  
 DB 72 QKFFKARENDPFCVSYPKTGTTWLTALTAFAANRKFEDVS-----TNPLKRNPHFEV 126  
 QY 138 LEYLNGRENTWPDML--NESPLFAGHIPYSILPASVLKSGTKIINISNRKSTTVSF 196  
 DB 127 IEIDF-----PPFSPVDVLKDEGNILFSTHPIYDILLPSVSKGKIYIWRDPKDTFVM 182  
 QY 197 WKFGNLIINPK--LLDLEKSVDFIFASGISFCGPEWNFQAEFTNAASTN--SNLLLSYEEM 253  
 DB 183 WFAFKERSQQGPVVSIEEAFKVCQGLSAYGPDVLDHVLGYKAYQANPDQILFLKYETM 242  
 QY 254 LEKPVNKKLAENFGCGFTDDEEKQGIIVDEIVKLCSPDNLKQVKNKNGSSYNSK---I 310  
 DB 243 RADPLPYKRLAENFGYFTKEEEEGNVVEKVKLCSPFETLNLEANKGEKDRDPVY 302  
 QY 311 DNKHFPRKGEVRDWNLYLTSEMIKKLETAGKINSEK--HLLSNKN 354  
 DB 303 ANSAVFRKGVGDWQNYLTPEWVARID--GLMEKFKGTGFLSSKS 346

RESULT 7  
 ABP72962  
 ID ABP72962 standard; protein; 338 AA.  
 AC ABP72962;  
 DT 28-MAY-2003 (first entry)  
 XX  
 DE Amino acid sequence of desulfoglucosinolate sulfotransferase AtST5a.  
 XX Plant; desulfoglucosinolate sulfotransferase; AtST5a; AtST5b; AtST5c;  
 KW glucosinolate; transgenic plant.  
 XX Arabidopsis thaliana.  
 OS Arabidopsis thaliana.  
 XX WO2003010318-A2.  
 PN  
 PD 06-FEB-2003.  
 XX  
 PF 24-JUL-2002; 2002WO-CA0001144.  
 XX  
 PR 24-JUL-2001; 2001US-0307141P.  
 XX  
 XX (UYCO-) UNIV CONCORDIA.  
 PA  
 XX Varin L, Spertini D;  
 XX  
 XX WPI; 2003-248082/24.  
 DR N-PSDB; ABZ68951.  
 XX  
 XX Novel isolated or purified polypeptide having biological activity of  
 PT desulfoglucosinolate sulfotransferases, useful as enzymes involved in  
 PT glucosinolate biosynthesis in transgenic plants.  
 XX  
 XX Claim 35; Page 56-57; 68pp; English.

XX The present sequence represents a plant desulfoglucosinolate  
 CC sulfotransferase, designated AtST5a. The specification also describe  
 CC AtST5b and AtST5c. AtST5a and AtST5b are constitutively expressed in all  
 CC parts of the plants, while AtST5c does not seem to be expressed in the  
 CC early stage of development. These enzymes are involved in glucosinolate  
 CC synthesis in plants. The desulfoglucosinolate sulfotransferase  
 CC polynucleotides are useful for modulating glucosinolate biosynthesis in  
 CC transgenic plants  
 XX Sequence 338 AA;  
 SQ  
 Query Match 27.3%; Score 562.5; DB 6; Length 338;  
 Best Local Similarity 38.4%; Pred. NO. 1.3e-40;  
 Matches 129; Conservative 59; Mismatches 123; Indels 25; Gaps 10;  
 QY 33 EKEEDSKMYKRYREIVSSLPNDY--GDTWRLYKGFQWQGYLVPGIMAFEDNFKARETD 90  
 DB 14 EUTEFEKTQKTKYQDFIATLPHKSGWRPDEILTYQGHWWQEBCLLEGLFHAKDHFARPTD 73  
 QY 91 IILTTLPAKATTTWKALTFAILTRDVNHPSPSTHPLFFNPHSCVQNLEYLYMGRENTMP 150  
 DB 74 FLVCSYPTKGTTLWKALTVAIVNR--SRVDDAANPLKKNPHEFVYVEIDFA---FYP 127  
 QY 151 DIDMLNESPR-LFAGHIPYSILPASVLKSGTKIINISNRKSTTVFVFWKGNLIINPK-- 207  
 DB 128 TVDVLQDRKNPLFSTHPIPNGLLPDSIVNSGCKWYIWRDPKDTFISMMTF---LHKEK 184  
 QY 208 ---LLDLEKSVDFIFASGISFCGPEWNFQAEFTNAASTN--SNLLLSYEEMLEKPVENYK 263  
 DB 185 EQGLASLEDSDFMFCGLSVGYDLDHVLGYKAYQENPDRIFLRYETMTMRANPLPFVKR 244  
 QY 264 LAEFMCGGTDEEKGQIVDEIVKLCSPDNLKQVKNKNGSSYNSK---IDNKHFRKGE 320  
 DB 245 LAEFMCGYGTDEEENGVAEKVKLCSPFETLNLEANKGDERDRPAPVANSAYFRK 304  
 QY 321 VRDWNLYLTSEMIKKLETAGKINSEK--HLLSNKN 354  
 DB 305 VGDWANYLTPEWVARID--GLVEEKFQDTGLLOHDN 338  
 RESULT 8  
 AAG16972  
 ID AAG16972 standard; protein; 338 AA.  
 AC AAG16972;  
 DT 17-OCT-2000 (first entry)  
 XX  
 DE Arabidopsis thaliana protein fragment SEQ ID NO: 17817.  
 XX Protein identification; signal transduction pathway; metabolic pathway;  
 KW hybridisation assay; genetic mapping; gene expression control; promoter;  
 KW termination sequence.  
 XX Arabidopsis thaliana.  
 OS  
 XX EP1033405-A2.  
 PN  
 PD 06-SEP-2000.  
 XX  
 XX 25-FEB-2000; 2000EP-00301439.  
 PF  
 XX 25-FEB-1999; 99US-0121825P.  
 PR 05-MAR-1999; 99US-0123180P.  
 PR 09-MAR-1999; 99US-0123548P.  
 PR 23-MAR-1999; 99US-0125788P.  
 PR 25-MAR-1999; 99US-0126264P.  
 PR 29-MAR-1999; 99US-0126785P.  
 PR 01-APR-1999; 99US-0127462P.  
 PR 06-APR-1999; 99US-0128234P.  
 PR 08-APR-1999; 99US-0128714P.  
 PR 16-APR-1999; 99US-0129845P.

PR 19-APR-1999; 99US-013007P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 18-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135134P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137232P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140343P.  
PR 24-JUN-1999; 99US-0140354P.  
PR 28-JUN-1999; 99US-0140695P.  
PR 29-JUN-1999; 99US-0140823P.  
PR 30-JUN-1999; 99US-0140991P.  
PR 01-JUL-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142300P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-APR-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 21-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 26-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145313P.  
PR 27-JUL-1999; 99US-0145318P.  
PR 27-JUL-1999; 99US-0145319P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 05-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 31-AUG-1999; 99US-0151303P.  
PR 01-SEP-1999; 99US-0151438P.  
PR 07-SEP-1999; 99US-0151930P.  
PR 10-SEP-1999; 99US-0152363P.  
PR 13-SEP-1999; 99US-0153070P.  
PR 15-SEP-1999; 99US-0153758P.  
PR 16-SEP-1999; 99US-0154018P.  
PR 20-SEP-1999; 99US-0154039P.  
PR 22-SEP-1999; 99US-0154779P.  
PR 23-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-015753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.





DT	31-MAY-2002	(first entry)	
XX	Herbicidally active polypeptide SEQ ID NO 800.		
XX	Herbicidal; plant; agriculture; herbicide.		
XX	Arabidopsis thaliana.		
XX	WO200210210-A2.		
XX	07-FEB-2002.		
XX	28-AUG-2001; 2001WO-EP009892.		
XX	28-AUG-2001; 2001WO-EP009892.		
XX	(FARB ) BAYER AG.		
PA	Tietjen K, Weidler M;		
PI	WPI; 2002-269010/31.		
XX			
XX	Identifying plant target proteins for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms.		
PT			
PT			
PT			
XX	Claim 5; SEQ ID NO 800; 261pp + Sequence Listing; English.		
PS			
XX	The invention relates to identifying target proteins (ABB90790-ABB94016) for herbicidally active compounds, comprising aligning and comparing nucleic acid or amino acid sequences from plant with nucleic acid or amino acid sequences from non-plant organisms using suitable search parameters, where plant sequences having an E-value greater by a factor of 3 than the E-value of most similar non-plant sequences are selected.		
CC	The polypeptides or nucleic acids encoding them are useful for identifying modulators. The identified modulators are useful as herbicides		
XX			
XX	Sequence 350 AA;		
XX	Query Match 27.1%; Score 558.5; DB 5; Length 350;		
XX	Best Local Similarity 39.0%; Pred. No. 3.1e-40;		
XX	Matches 129; Conservative 59; Mismatches 120; Indels 23; Gaps 9		
Qy	29 KNEQKEEDSKMKYREIVKSLSDNYAGDTMFL--YKGFQWQGVLPQIMAFEDNKA 86		
Db	22 ETKTSTEFKQKRYQDLISTFPEHKGWRPKPLEIYGQWMLPSLLEGCIHAQEFFQA 81		
Qy	87 RETDIIITLTPKAGTTWTKALTFTAILTRDYNHPSPTHLLFFNPHSCVQNLEYLYWGR 146		
Db	82 RPSDFLVCSYPTGTTLTKALTFANRSRFDSS--NPLLKRNPHEFVPYIEDF--- 135		
Qy	147 NTPDIDLMLNESPR-LFAGHIPIYSLPASVLKSGTKIINSRNKKSTFVSFWKGNLNP 205		
Db	136 PFFPEVDNLKQNTLFTSHIPIYELLPSVSVKSGCKMYIWRPEKDTFISMTTF---LHK 192		
Qy	206 DK-----LLDLEKSVDIFASGTSFGCPWFNFQAEFTNAASTN-SNLLLSYEEMLEKPYE 259		
Db	193 ERTELGPVSNLEBSDFMFCRGLSGYGYPLNHLAYKRAYQENPDRIULFKYETMRADPLP 252		
Qy	260 NVKLAIEFGCGFTDEEEKQGVLDIVKLCSDNLKQCVNKGSSVNSK---IDNKHFF 316		
Db	253 YVKSIAEFFGHGFTAEEEEKGVVEKVNLCSPFELLNLANKGKEDREDPGVANSAYF 312		
Qy	317 RKGEVDRWANYLTSETMIKKLETAGKINSEK 347		
Db	313 RKGKVGDSNYLTPEMAAID--GLMEKFK 341		

RESULT 11  
ABP72963  
ID ABP7

XX	ABP72963;
AC	
CC	28-MAY-2003 (first entry)
DE	Amino acid sequence of desulfoglucosinolate sulfotransferase AtST5b.
KW	Plant; desulfoglucosinolate sulfotransferase; AtST5a; AtST5b; AtST5c; glucosinolate; transgenic plant.
OS	Arabidopsis thaliana.
PN	WO2003010318-A2.
PD	06-FEB-2003.
PX	24-JUL-2002; 2002WO-CA001144.
PR	24-JUL-2001; 2001US-0307141P.
PA	(UYCO-) UNIV CONCORDIA.
PI	Varin L, Spertini D;
PP	WPI: 2003-248082/24.
DR	N-PSDB; ABZ68952.
CC	Novel isolated or purified polypeptide having biological activity of desulfoglucosinolate sulfotransferases, useful as enzymes involved in glucosinolate biosynthesis in transgenic plants.
XX	Claim 35; Page 59-60; 68pp; English.
CC	The present sequence represents a plant desulfoglucosinolate sulfotransferase, designated AtST5b. The specification also describes AtST5a and AtST5c. AtST5a and AtST5b are constitutively expressed in all parts of the plants, while AtST5c does not seem to be expressed in the early stage of development. These enzymes are involved in glucosinolate synthesis in plants. The desulfoglucosinolate sulfotransferase polynucleotides are useful for modulating glucosinolate biosynthesis in transgenic plants
SQ	Sequence 350 AA;
QY	Query Match 27.1%; Score 558.5; DB 6; Length 350;
Db	Best Local Similarity 39.0%; Pred. No. 3.le-40;
QY	Matches 129; Conservative 59; Mismatches 120; Indels 23; Gaps 9
Db	29 KNEQEKEEDSKMYKRYEIRVSSLSFSDNDVWGDTMEL--YKGFWQMGYLVPGIMAFEDNPKA 86
QY	22 EYKTESEFEKNQKYQLDITFFPHKGRPKPELIEYGCVWLPSLLEGCIIHAQEFFQA 81
Db	87 RETDIILTLTPKAGTTWTKALTFAILLTRDVNHSPSPTPLLFNPHSVCQNLEYLYMGRE 146
Db	82 RPSDFVLCVSYPKTGTITLKALTFAIANRSRDDS--NPLIKRNPFHFVFYIETD---- 135
QY	147 NTPMDLDMLNESPR-LFAGHTPYSLLPASVLKSGTKINISNRKRSTFVSFWFGNLINP 205
Db	136 PFPPEVDVLKDGNVLFSTHPIYELLPSVSVYKSCKMVYWRSPKOTFISMWTF---LHK 192
QY	206 DK-----LLDLSEKVDIFASGISPCGPEWNFAQFTNAASTN-SNILLSYEEMLEKPYE 259
Db	193 ERTELGVSNLEESDFMFCRGSLGGVGYLNHILAYWKAYOENFDRIFLFKYETMRADPIP 252
QY	260 NVVKLAEEFGCGFTDEEKQGIVDREVKLCSFDNLKKQQVNKGSSYSNK---IDNKHFF 316
Db	253 YVKSLEAFFGHGFTAEEBEKGVVEKVNVNLCSEFETLKNLEANKGEKREDRPGVYANSAYF 312
QY	317 RKGEVRDWANYLTSEMINKLETAGKINESEK 347
Db	313 RKGKVGDSNYLTPPENAARID--GLMEKF 341

RESULT 12  
AAG46836  
ID AAG46836 standard; protein; 359 AA.  
XX AC AAG46836;  
XX DT 18-OCT-2000 (first entry)  
XX DE Arabidopsis thaliana protein fragment SEQ ID NO: 58965.  
XX KW Protein identification; signal transduction pathway; metabolic pathway;  
KW hybridisation assay; genetic mapping; gene expression control; promoter;  
KW termination sequence.  
XX OS Arabidopsis thaliana.  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130445P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131049P.  
PR 30-APR-1999; 99US-0132048P.  
PR 04-MAY-1999; 99US-0132407P.  
PR 05-MAY-1999; 99US-0132484P.  
PR 06-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 18-MAY-1999; 99US-0134370P.  
PR 19-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134944P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 22-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145090P.  
PR 22-JUL-1999; 99US-0145152P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 09-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149436P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.



PR	20-AUG-1999;	99US-0149929P.	
PR	23-AUG-1999;	99US-0149902P.	
PR	23-AUG-1999;	99US-0149930P.	
PR	25-AUG-1999;	99US-0150566P.	
PR	26-AUG-1999;	99US-0150884P.	
PR	27-AUG-1999;	99US-0151065P.	
PR	27-AUG-1999;	99US-0151066P.	
PR	30-AUG-1999;	99US-0151080P.	
PR	30-AUG-1999;	99US-0151303P.	
PR	31-AUG-1999;	99US-0151438P.	
PR	01-SEP-1999;	99US-0151930P.	
PR	07-SEP-1999;	99US-0152363P.	
PR	10-SEP-1999;	99US-0153070P.	
PR	13-SEP-1999;	99US-0153758P.	
PR	16-SEP-1999;	99US-0154018P.	
PR	20-SEP-1999;	99US-0154039P.	
PR	20-SEP-1999;	99US-0154779P.	
PR	22-SEP-1999;	99US-0155139P.	
PR	23-SEP-1999;	99US-0155486P.	
PR	24-SEP-1999;	99US-0155659P.	
PR	28-SEP-1999;	99US-0156458P.	
PR	29-SEP-1999;	99US-0156596P.	
PR	04-OCT-1999;	99US-0157117P.	
PR	05-OCT-1999;	99US-0157753P.	
PR	06-OCT-1999;	99US-0157865P.	
PR	07-OCT-1999;	99US-0158029P.	
PR	08-OCT-1999;	99US-0158232P.	
PR	12-OCT-1999;	99US-0158369P.	
PR	13-OCT-1999;	99US-0159293P.	
PR	13-OCT-1999;	99US-0159294P.	
PR	13-OCT-1999;	99US-0159295P.	
PR	14-OCT-1999;	99US-0159329P.	
PR	14-OCT-1999;	99US-0159330P.	
PR	14-OCT-1999;	99US-0159331P.	
PR	14-OCT-1999;	99US-0159637P.	
PR	18-OCT-1999;	99US-0159638P.	
PR	21-OCT-1999;	99US-0159584P.	
PR	21-OCT-1999;	99US-0160741P.	
PR	21-OCT-1999;	99US-0160767P.	
PR	21-OCT-1999;	99US-0160768P.	
PR	21-OCT-1999;	99US-0160770P.	
PR	21-OCT-1999;	99US-0160814P.	
PR	22-OCT-1999;	99US-0160980P.	
PR	22-OCT-1999;	99US-0160981P.	
PR	22-OCT-1999;	99US-0160989P.	
PR	25-OCT-1999;	99US-0161404P.	
PR	25-OCT-1999;	99US-0161405P.	
PR	25-OCT-1999;	99US-0161406P.	
PR	26-OCT-1999;	99US-0161359P.	
PR	26-OCT-1999;	99US-0161360P.	
PR	28-OCT-1999;	99US-0161361P.	
PR	28-OCT-1999;	99US-0161920P.	
PR	28-OCT-1999;	99US-0161992P.	
PR	28-OCT-1999;	99US-0161993P.	
PR	29-OCT-1999;	99US-0162142P.	

Query Match

Best Local Similarity 27.0%; Score 555.5; DB 3; Length 359;

Matches 123; Conservative 69; Mismatches 140; Indels 11; Gaps 8;

QY	1	TRGITGIVACUPLINAGTIALEKFCGKNEQKEEDSKMYKRYREIVSSLPNSNDYGD	60
DB	3	TSSMKSIPIAIFSFMSCHLLEKKEKTRDVPKABEEDGLSCFQEMLDLSLPKRGWTR	62
QY	61	-VRLYKGFQMGVLPVGINAFEDNFKARETDIITLPKAGTTWTKALTFAITLTDVNH	119
DB	63	YLYLFGFGWCQAKETQAIMSKFKHOSLENDVVLATIPKSGTWNKALTFTILNHRDP	122
QY	120	--SSPTHTPLFFNPHSCVQNLLEY-LYMGRENTWPDMLNESPRLFAGHIPYSLLPASVL	176
DB	123	VASSTNHLFTSNPHDLVPFFVEYKLYANGD--VPDLGL-ASPRTFATHLPFGSLKMTIE	179

QY	177	KSGTKIINISNRKSTFVSFWKFGNLINPDKL--LDLEKSVDFIFASGISFCQPEWFOAE	234
DB	180	KPGVKVYLCRNPFDTFISSWHYTNNIKSESVPVLLDQAFDLYCRGVIGFGFWEHMLG	239
QY	235	P-TNAASTNSNLLLSYEMLEKPEVNVKKLAEFMGCCTDDEEKOGIVDEIVKLCSPDN	293
DB	240	YWRESLKRPEKVFYFIEDLKDDITNLRKLTATFLELFTFEERKGVVKAIAELCSPEN	299
QY	294	LKNQOVNKGSSSYNSKIDNKHFFRKGEVRDWNAYILTSEMIKKL	336
DB	300	LKKLEVKNKSIKN-FENRFLRKGEVSDWVNYLSPSQVERL	341

RESULT 13

AAG44205

ID AAG44205 standard; protein; 359 AA.

XX AAG44205;

XX

DT 18-OCT-2000 (first entry)

XX

DE Arabidopsis thaliana protein fragment SEQ ID NO: 55343.

XX

KW Protein identification; signal transduction pathway; metabolic pathway; hybridisation assay; genetic mapping; gene expression control; promoter; termination sequence.

KW

XX Arabidopsis thaliana.

XX

XX EP1033405-A2.

XX

PD 06-SEP-2000.

XX

EF 25-FEB-2000; 2000EP-00301439.

XX

PR 25-FEB-1999; 99US-0121825P.

PR

PR 05-MAR-1999; 99US-0123180P.

PR

PR 09-MAR-1999; 99US-0123548P.

PR

PR 23-MAR-1999; 99US-0125788P.

PR

PR 28-MAR-1999; 99US-0126264P.

PR

PR 29-MAR-1999; 99US-0126785P.

PR

PR 01-APR-1999; 99US-0127462P.

PR

PR 06-APR-1999; 99US-0128234P.

PR

PR 08-APR-1999; 99US-0128714P.

PR

PR 16-APR-1999; 99US-0129845P.

PR

PR 19-APR-1999; 99US-0130077P.

PR

PR 21-APR-1999; 99US-0130449P.

PR

PR 23-APR-1999; 99US-0130510P.

PR

PR 23-APR-1999; 99US-0130891P.

PR

PR 28-APR-1999; 99US-0131449P.

PR

PR 30-APR-1999; 99US-0132048P.

PR

PR 30-APR-1999; 99US-0132407P.

PR

PR 04-MAY-1999; 99US-0132484P.

PR

PR 05-MAY-1999; 99US-0132485P.

PR

PR 06-MAY-1999; 99US-0132486P.

PR

PR 06-MAY-1999; 99US-0132487P.

PR

PR 07-MAY-1999; 99US-0132863P.

PR

PR 11-MAY-1999; 99US-0134256P.

PR

PR 14-MAY-1999; 99US-0134218P.

PR

PR 14-MAY-1999; 99US-0134219P.

PR

PR 14-MAY-1999; 99US-0134221P.

PR

PR 14-MAY-1999; 99US-0134370P.

PR

PR 18-MAY-1999; 99US-0134768P.

PR

PR 19-MAY-1999; 99US-0134941P.

PR

PR 20-MAY-1999; 99US-0135124P.

PR

PR 21-MAY-1999; 99US-0135353P.

PR

PR 24-MAY-1999; 99US-0135629P.

PR

PR 25-MAY-1999; 99US-0136021P.

PR

PR 27-MAY-1999; 99US-0136392P.

PR

PR 28-MAY-1999; 99US-0136782P.

PR

PR 01-JUN-1999; 99US-0137222P.

PR

PR 03-JUN-1999; 99US-0137528P.

PR

PR 04-JUN-1999; 99US-0137502P.

PR





PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132862P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137528P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139453P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140931P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0130891P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145313P.  
PR 27-JUL-1999; 99US-0145318P.  
PR 27-JUL-1999; 99US-0145319P.  
PR 28-JUL-1999; 99US-0145351P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 06-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 09-AUG-1999; 99US-0148171P.  
PR 10-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 17-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149923P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 30-AUG-1999; 99US-0151080P.  
PR 31-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157553P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158023P.  
PR 08-OCT-1999; 99US-0158234P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159332P.  
PR 14-OCT-1999; 99US-0159337P.

```
PR 14-OCT-1999; 99US-0159638P.
PR 18-OCT-1999; 99US-0159842P.
PR 21-OCT-1999; 99US-0160741P.
PR 21-OCT-1999; 99US-0160767P.
PR 21-OCT-1999; 99US-0160768P.
PR 21-OCT-1999; 99US-0160770P.
PR 21-OCT-1999; 99US-0160814P.
PR 21-OCT-1999; 99US-0160815P.
PR 22-OCT-1999; 99US-0160980P.
PR 22-OCT-1999; 99US-0160981P.
PR 25-OCT-1999; 99US-0160989P.
PR 25-OCT-1999; 99US-0161404P.
PR 25-OCT-1999; 99US-0161405P.
PR 25-OCT-1999; 99US-0161408P.
PR 26-OCT-1999; 99US-0161359P.
PR 26-OCT-1999; 99US-0161360P.
PR 28-OCT-1999; 99US-0161361P.
PR 28-OCT-1999; 99US-0161920P.
PR 28-OCT-1999; 99US-0161992P.
PR 28-OCT-1999; 99US-0161993P.
PR 29-OCT-1999; 99US-0162142P.

Query Match 26.8%; Score 551.5; DB 3; Length 354;
Best Local Similarity 36.2%; Pred. No. 1.3e-39;
Matches 122; Conservative 68; Mismatches 136; Indels 11; Gaps 8;

QY 7 IAVACPLTMAGILALEKCFGSKNEKEEDSMYKRYREIVSSLPSPNDYMGDT-MRLYK 65
Db 4 IPWAIPSPFNCHKLELLKEGKTRDVPKAEDEGLSCFEQEMLDSPKRGWTRYLILFQ 63

QY 66 GFWMQGLVFGIMAFEDNFKARETDIILTLPKAGTTWKALTFAILTRDVNHP--SSPT 123
Db 64 GFWCQAKEIQAIMSFQKHFQSLNDWLATIPKSGTTWLKALTFTILNRHRFPVASSTN 123

QY 124 HELLFNPSPCVQNLLEY-LYMGRENTMPDLDMLNESPRLFAGHIPYSLLPASVLKSGTKI 182
Db 124 HELFTSNPHDLVPFFFEYKLYANGD--VPDLGL-ASPRTFATHLPFGSLKETIERPGVKV 180

QY 183 INISNRKSTFVSFMKFGNLINPKL--LDLEKSVDFI PASGISFCGPENWFOAEF-TNAA 239
Db 181 VVLCRNPFDTFSSWHYTNNIKSESVPVLLDQAFDLYCRGVIGFGPFWEHMLGYWRESL 240

QY 240 STNSNLLLSYBEMLEKPVENVKLAEFNGCGFTDDEEKQGI VDEIVKLCSPDNLKNOQV 299
Db 241 KRPEKVFFLRVEDLKDDIETNLKRLATFLFLELPFTTEEEKRGVVKAIABLCSPENLKUEV 300

QY 300 NKNGSSYNGKIDNKHFFRKGEYRDWANYLTSEMIKKL 336
Db 301 NKNKSIKN-FENRFLFRKGEVSDWNYLSPSQVERL 336
```

Search completed: March 20, 2004, 01:37:46  
Job time : 49 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:36:08 ; Search time 17 Seconds  
(without alignments)  
1187.397 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059  
Sequence: 1 TRGITRAVACPLIMAGIL.....CLLSYFVINPILKKKK 391

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2\_6/ptodata/2/iaa/5A COMB.pdp.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B COMB.pdp.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A COMB.pdp.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B COMB.pdp.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCRTS COMB.pdp.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pdp.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	268	13.0	304	4	US-09-609-816-6
2	264	12.8	304	4	US-09-609-816-4
3	260	12.6	304	4	US-09-609-816-5
4	255.5	12.4	283	4	US-09-609-816-8
5	255.5	12.4	283	4	US-09-609-816-9
6	255	12.4	294	1	US-08-325-562-2
7	255	12.4	294	1	US-08-437-795-2
8	246.5	12.0	283	4	US-09-609-816-10
9	245.5	11.9	295	3	US-09-150-133-13
10	245.5	11.9	295	3	US-09-374-493-13
11	245.5	11.9	295	3	US-09-374-824-13
12	245.5	11.9	295	3	US-09-374-492-13
13	245.5	11.9	295	3	US-09-785-343-13
14	245.5	11.6	265	4	US-09-795-926-6
15	239	11.6	303	4	US-09-795-926-2
16	239	11.6	303	4	US-09-795-926-3
17	226.5	11.0	295	4	US-09-328-174A-3
18	199.5	9.7	350	2	US-08-852-481-2
19	187	9.1	284	4	US-09-786-240-7
20	184	8.9	214	4	US-09-795-926-16
21	155.5	7.6	148	4	US-09-795-926-8
22	155.5	7.6	148	4	US-09-795-926-12
23	111	5.4	110	4	US-09-795-926-4
24	111	5.4	148	4	US-09-795-926-10
25	111	5.4	1147	3	US-08-470-260-5
26	111	5.4	1147	3	US-08-471-491-5
27	111	5.4	1147	3	US-08-466-662-5

28 111 5.4 3289 2 US-08-477-451-2 Sequence 2, Appli  
29 103 5.0 859 1 US-08-053-614-2 Sequence 2, Appli  
30 103 5.0 859 1 US-08-316-397B-2 Sequence 2, Appli  
31 103 5.0 859 2 US-09-034-306-2 Sequence 2, Appli  
32 103 5.0 859 3 US-09-259-437-2 Sequence 2, Appli  
33 103 5.0 859 5 PCT-US93-09782-2 Sequence 2, Appli  
34 103 5.0 1181 1 US-08-053-614-4 Sequence 4, Appli  
35 103 5.0 1181 1 US-08-316-397B-4 Sequence 4, Appli  
36 103 5.0 1181 2 US-09-034-306-4 Sequence 4, Appli  
37 103 5.0 1181 3 US-09-259-437-4 Sequence 4, Appli  
38 103 5.0 1181 5 PCT-US93-09782-4 Sequence 4, Appli  
39 103 5.0 1467 4 US-09-134-000C-6740 Sequence 6740, Ap  
40 100.5 4.9 97 4 US-09-795-926-18 Sequence 18, Appli  
41 99.5 4.8 717 4 US-09-509-800-2 Sequence 2, Appli  
42 99 4.8 389 4 US-09-543-681A-7496 Sequence 7496, Ap  
43 98.5 4.8 1093 3 US-09-315-793-52 Sequence 52, Appli  
44 98 4.8 369 4 US-09-107-532A-5754 Sequence 5754, Ap  
45 97.5 4.7 840 3 US-08-974-549A-190 Sequence 190, App

#### ALIGNMENTS

##### RESULT 1

US-09-609-816-6  
; Sequence 6, Application US/09609816  
; Patent No. 6436684  
; GENERAL INFORMATION:  
; APPLICANT: Woodage, Trevor  
; APPLICANT: Wei, Minh Hui  
; APPLICANT: Kodira, Chinappa  
; APPLICANT: Beasley, Ellen  
; APPLICANT: DiFrancesco, Valentina  
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING  
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN  
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF  
; FILE REFERENCE: CL000669PCT  
; CURRENT APPLICATION NUMBER: US/09/609,816  
; CURRENT FILING DATE: 2000-07-03  
; PRIOR APPLICATION NUMBER: 60/192,408  
; PRIOR FILING DATE: 2000-03-27  
; PRIOR APPLICATION NUMBER: 60/212,725  
; PRIOR FILING DATE: 2000-06-20  
; PRIOR APPLICATION NUMBER: 09/609,816  
; PRIOR FILING DATE: 2000-07-03  
; NUMBER OF SEQ ID NOS: 23  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 304  
; TYPE: PRT  
; ORGANISM: HUMAN  
US-09-609-816-6

Query Match 13.0%; Score 268; DB 4; Length 304;  
Best Local Similarity 27.9%; Pred. No. 3.5e-17;  
Matches 97; Conservative 56; Mismatches 131; Indels 64; Gaps 18;  
QY 29 KNEQKEESKMKYREIVTSLSPSNDYWDTRLYKGFMQGYLVPGIMAFEDNPKARE 88  
Db 6 KNAPTWEKPELFINME--VDGVP-----TLISKWEWEK-----VCNFOAKP 46  
QY 89 TDILITLTKAGTWTWKALTFAILT-RDYNH----PSSETHPLFFN-PHSCVQNLLEYL 142  
Db 47 DDILILATYKSGITWHEILDMLNDGDYKCKRAQTLORHAFLEKLPFKKEKPDLEFV- 105  
QY 143 MGRNTWPDMLNESPRLPAGHIPVSLPASVLKSGTKIINISNRKSTFFVSWFKGNL 202  
Db 106 -----LEM--SSPOLIKTHLPGLIPPGIKENCKIVYVARNPKDCLVSYHFHMR 154  
QY 203 IN--PKKLIDLEKSVDFISAGISFCFPEWFOAEFTNAASTNSNLLLSYEENLEKPVEN 260  
Db 155 ASFWPDP-QNLEEFYKFMSC-KVVGSGFDHVKGWAAKDRHILLYFYEDIKNPKHE 212

```
QY 261 VKKLAFFMCGGTTDEEKQGIYDEIVKLCSPDNKQVKNQNGSSYNSKIDN---KHFFR 317
Db 213 IHKVLFFLEKTSUGD-----VINKLVHHTSFDMKDNPM-AHNTAVPAHFNHSISKEMR 266
QY 318 KGEVRDWANYLTSEMIMKLETAGKINES-EKHLNKNLNTFNLNYYC 364
Db 267 KGMFGDWKNHFTVAM-----NENFDKHY--EKKMAGSTL-NFC 301

RESULT 2
US-09-609-816-4
; Sequence 4, Application US/09609816
; Patent No. 6436684
; GENERAL INFORMATION:
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: DiFrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; TYPE: PRT
; ORGANISM: HUMAN
; US-09-609-816-4

Query Match 12.8%; Score 264; DB 4; Length 304;
Best Local Similarity 27.6%; Pred. No. 8.3e-17;
Matches 96; Conservative 58; Mismatches 130; Indels 64; Gaps 18;

QY 29 KNEQKEEDSKMYKRYREIVSSLPNSNDYWGDTWRLYKGFQWQGYLVPGIMAFEDNFKARE 88
Db 6 KNAFTMEKKPELFNIME--VDGVP-----TLILSKEWWEK-----VCNFQAKP 46
QY 89 TDILTLTPKAGTTWTKALTFAILT-RDVNH-----PSSPHTPLIFFN-PHSCVQNLEYLY 142
Db 47 DDLILATVPKSGTTWHTHEILDMLNDGDEVKCKRAQTLDRHAFLEKFPKKEPDLEFV- 105
QY 143 MGRNTMPDLDMLNESPRLPAGHIPYSLLPASVLKSGTKIINISRNKRKSTFVSFWKFGNL 202
Db 106 -----LEM--SSQLIKTHLPShLIPPSIWKENCKIYVYARNPKDCLVSYHFFRM 154
QY 203 IN--PKLLDLEKSDVDFASGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLEKPVEN 260
Db 155 ASFPMDP-QNLEEFYEFKFMG-KVVGGSWFDHVKGWAAKDMHRILYLFYEDIKKDPKRE 212
QY 261 VKKLAFFMCGGTTDEEKQGIYDEIVKLCSPDNKQVKNQNGSSYNSKIDN---KHFFR 317
Db 213 IHKVLFFLEKTSUGD-----VINKLVHHTSFDMKDNPM-AHNTAVPAHFNHSISKEMR 266
QY 318 KGEVRDWANYLTSEMIMKLETAGKINES-EKHLNKNLNTFNLNYYC 364
Db 267 KGMFGDWKNHFTVAM-----NENFDKHY--EKKMAGSTL-NFC 301

RESULT 3
US-09-609-816-5
; Sequence 5, Application US/09609816
; Patent No. 6436684
; GENERAL INFORMATION:
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: DiFrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4
; TYPE: PRT
; ORGANISM: HUMAN
; US-09-609-816-5

Query Match 12.8%; Score 264; DB 4; Length 304;
Best Local Similarity 27.6%; Pred. No. 8.3e-17;
Matches 96; Conservative 58; Mismatches 130; Indels 64; Gaps 18;

QY 29 KNEQKEEDSKMYKRYREIVSSLPNSNDYWGDTWRLYKGFQWQGYLVPGIMAFEDNFKARE 88
Db 6 KNAFTMEKKPELFNIME--VDGVP-----TLILSKEWWEK-----VCNFQAKP 46
QY 89 TDILTLTPKAGTTWTKALTFAILT-RDVNH-----PSSPHTPLIFFN-PHSCVQNLEYLY 142
Db 47 DDLILATVPKSGTTWHTHEILDMLNDGDEVKCKRAQTLDRHAFLEKFPKKEPDLEFV- 105
QY 143 MGRNTMPDLDMLNESPRLPAGHIPYSLLPASVLKSGTKIINISRNKRKSTFVSFWKFGNL 202
Db 106 -----LEM--SSQLIKTHLPShLIPPSIWKENCKIYVYARNPKDCLVSYHFFRM 154
QY 203 IN--PKLLDLEKSDVDFASGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLEKPVEN 260
Db 155 ASFPMDP-QNLEEFYEFKFMG-KVVGGSWFDHVKGWAAKDMHRILYLFYEDIKKDPKRE 212
QY 261 VKKLAFFMCGGTTDEEKQGIYDEIVKLCSPDNKQVKNQNGSSYNSKIDN---KHFFR 317
Db 213 IHKVLFFLEKTSUGD-----VINKLVHHTSFDMKDNPM-AHNTAVPAHFNHSISKEMR 266
QY 318 KGEVRDWANYLTSEMIMKLETAGKINES-EKHLNKNLNTFNLNYYC 364
Db 267 KGMFGDWKNHFTVAM-----NENFDKHY--EKKMAGSTL-NFC 301
```

```
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: DiFrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 304
; TYPE: PRT
; ORGANISM: HUMAN
; US-09-609-816-5

Query Match 12.6%; Score 260; DB 4; Length 304;
Best Local Similarity 27.7%; Pred. No. 2e-16;
Matches 86; Conservative 54; Mismatches 123; Indels 48; Gaps 13;

QY 29 KNEQKEEDSKMYKRYREIVSSLPNSNDYWGDTWRLYKGFQWQGYLVPGIMAFEDNFKARE 88
Db 6 KNAFTMEKKPELFNIME--VDGVP-----TLILSKEWWEK-----VCNFQAKP 46
QY 89 TDILTLTPKAGTTWTKALTFAILT-RDVNH-----PSSPHTPLIFFN-PHSCVQNLEYLY 142
Db 47 DDLILATVPKSGTTWHTHEILDMLNDGDEVKCKRAQTLDRHAFLEKFPKKEPDLEFV- 105
QY 143 MGRNTMPDLDMLNESPRLPAGHIPYSLLPASVLKSGTKIINISRNKRKSTFVSFWKFGNL 202
Db 106 -----LEM--SSQLIKTHLPShLIPPSIWKENCKIYVYARNPKDCLVSYHFFRM 154
QY 203 IN--PKLLDLEKSDVDFASGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLEKPVEN 260
Db 155 ASFPMDP-QNLEEFYEFKFMG-KVVGGSWFDHVKGWAAKDMHRILYLFYEDIKKDPKRE 212
QY 261 VKKLAFFMCGGTTDEEKQGIYDEIVKLCSPDNKQVKNQNGSSYNSKIDN---KHFFR 318
Db 213 IHKVLFFLEKTSUGD-----VINKLVHHTSFDMKDNPM-AHNTAVPAHFNHSISKEMR 267
QY 319 GEVRDWANYLT 329
Db 268 GMPGDWKNYFT 278

RESULT 4
US-09-609-816-8
; Sequence 8, Application US/09609816
; Patent No. 6436684
; GENERAL INFORMATION:
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: DiFrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
```

```
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 283
; TYPE: PRT
; ORGANISM: HUMAN
US-09-609-816-8

Query Match
Best Local Similarity 12.4%; Score 255.5; DB 4; Length 283;
Matches 76; Conservative 56; Mismatches 109; Indels 51; Gaps 10;

QY 83 NFKARETDIIILTLPLKAGTTWTAKLTFAI-----LTRDVNHPSPSPHLLFPNPHSCVQN 137
Db 20 NFOAKPDLLIATYAKAGTTWTQEIIVDMIONDGVQKQORANTYDRHPFI-----69

QY 138 LEYLVMGRENTPD-----LDMLE--SPRLFAGHIPYSLLPASVLKSGTKIINISNRK 190
Db 70 -----EWTLPSPNSGLDLANKMPSPTLTKTLPHVHMLPSPFWKENSIIYVARNK 121

QY 191 STFVSFWKGNLNPDKL-----DLKSYDIFASGISFCGPEWNFQAEFTNAASTNSLL 246
Db 122 DCLVSYIYFSRM---NKMPLDPDTGLGEYIEQFAGKVLWG-SWYDHVKGWWDVKQHRIL 177

QY 247 LLSYEEMLEKPVENVKLAEFMCGFTDDEEKQGIYDEIVKLCSPDNLKQVKNKNGSSY 306
Db 178 YLFYEDMKEDPKREIKKIAKFL-----EKDISSEVLNKIIYHTSFVDMKENPMANYTTLP 232

QY 307 NSKIDN--KHFFKGEVRDWNALYLTSEMICKLETAGKINSEKHLSSKNLT 356
Db 233 SSIMDHSISPFMRKMGPGDWKNYFT-----VAQSEDFDEYRRKAGSNIT 278

RESULT 5
US-09-609-816-9
; Sequence 9, Application US/09609816
; Patent No. 643684
; GENERAL INFORMATION:
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: DiFrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 283
; TYPE: PRT
; ORGANISM: HUMAN
US-09-609-816-9

Query Match
Best Local Similarity 12.4%; Score 255.5; DB 4; Length 283;
Matches 76; Conservative 56; Mismatches 109; Indels 51; Gaps 10;

QY 83 NFKARETDIIILTLPLKAGTTWTAKLTFAI-----LTRDVNHPSPSPHLLFPNPHSCVQN 137
Db 20 NFOAKPDLLIATYAKAGTTWTQEIIVDMIONDGVQKQORANTYDRHPFI-----69

QY 138 LEYLVMGRENTPD-----LDMLE--SPRLFAGHIPYSLLPASVLKSGTKIINISNRK 190
Db 70 -----EWTLPSPNSGLDLANKMPSPTLTKTLPHVHMLPSPFWKENSIIYVARNK 121

QY 191 STFVSFWKGNLNPDKL-----DLKSYDIFASGISFCGPEWNFQAEFTNAASTNSLL 246
Db 122 DCLVSYIYFSRM---NKMPLDPDTGLGEYIEQFAGKVLWG-SWYDHVKGWWDVKQHRIL 177

QY 247 LLSYEEMLEKPVENVKLAEFMCGFTDDEEKQGIYDEIVKLCSPDNLKQVKNKNGSSY 306
Db 178 YLFYEDMKEDPKREIKKIAKFL-----EKDISSEVLNKIIYHTSFVDMKENPMANYTTLP 232

QY 307 NSKIDN--KHFFKGEVRDWNALYLTSEMICKLETAGKINSEKHLSSKNLT 356
Db 233 SSIMDHSISPFMRKMGPGDWKNYFT-----VAQSEDFDEYRRKAGSNIT 278

US-08-325-562-2
; Sequence 2, Application US/08325562
; Patent No. 5714594
; GENERAL INFORMATION:
; APPLICANT: Weinshilboum, Richard M.
; APPLICANT: Aksoy, Ibrahim A.
; APPLICANT: Wood, Thomas C.
; TITLE OF INVENTION: CDNA CLONING AND EXPRESSION OF HUMAN
; TITLE OF INVENTION: LIVER ESTROGEN SULFOESTERASE
; NUMBER OF SEQUENCES: 17
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/325,562
; FILING DATE: 18-OCT-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Muetting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 150.131US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 294 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-325-562-2

Query Match
Best Local Similarity 12.4%; Score 255; DB 1; Length 294;
Matches 85; Conservative 51; Mismatches 120; Indels 54; Gaps 13;

QY 41 YKRYEIVSSLPSPNDYWGDTWRLYKGFQMGVLYPGIMAFEDN---FKARETDIIILTLPL 97
Db 8 YEKPEVHGIL-----MYKDF-----VKYWDNVFAFOARDDLVIAIYP 46

QY 98 KAGTTWTAKLTFAI--LVNHPSPSPHLLFPNPHSCVQNLEYLVGMRENTMPDMLN 156
Db 47 KSGTTWSEIVVMYKGEDEVKCKED-----VIFN-----RIPFLECKENLMNGVKQLD 96

QY 157 E--SPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFVSFWKGNLI-----NPDKLLD 210
```



Db	97	EWNSRIVKTHLPPELLPASTWEKCDKIIILCRNAKDVAVSFYFFLWAGHPGSPFPE	156
QY	211	LEKSYDIFASGISFCGPPWNPQAFETNAASTNSLLLSYEEMLEKPEVNVKLAEFMGC	270
Db	157	F---VEKEMQGVPG-SWYKHVKSWEKGSPEVLFLEYEDLKEDIRKEV-KLIHFL--	210
QY	271	GFTDDEEQGIVDEIVKLCSPDNLKNQVKNKGSYNSKIDNKH--FPRKGEVRDWAY	327
Db	211	---ERKPSBELVDRIIHTSFOEMKNKP-SINYITLPDEIMNQKLSFFMKRKGITGDWKH	266
QY	328	LTSEMIKKIE	337
Db	267	FTVALNEXFD	276

RESULT 7  
 US-08-437-795-2  
 ; Sequence 2, Application US/08437795  
 ; Patent No. 5744355  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Weinshilboum, Richard M.  
 ; APPLICANT: Aksoy, Ibrahim A.  
 ; APPLICANT: Wood, Thomas C.  
 ; TITLE OF INVENTION: CDNA CLONING AND EXPRESSION OF HUMAN  
 ; TITLE OF INVENTION: LIVER ESTROGEN SULFOTRANSFERASE  
 ; NUMBER OF SEQUENCES: 19  
 ; CORRESPONDENCE ADDRESS:  
 ; ADDRESSEE: Schwegman, Lundberg & Woessner P.A.  
 ; STREET: 3500 IDS Center  
 ; CITY: Minneapolis  
 ; STATE: MN  
 ; COUNTRY: USA  
 ; ZIP: 55402  
 ; COMPUTER READABLE FORM:  
 ; MEDIUM TYPE: Floppy disk  
 ; COMPUTER: IBM PC Compatible  
 ; OPERATING SYSTEM: PC-DOS/MS-DOS  
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25  
 ; CURRENT APPLICATION DATA:  
 ; APPLICATION NUMBER: US/08/437,795  
 ; FILING DATE: 09-MAY-1995  
 ; CLASSIFICATION: 435  
 ; ATTORNEY/AGENT INFORMATION:  
 ; NAME: Muetting, Ann M.  
 ; REGISTRATION NUMBER: 33,977  
 ; REFERENCE/DOCKET NUMBER: 150.137US1  
 ; TELECOMMUNICATION INFORMATION:  
 ; TELEPHONE: 612-339-0331  
 ; TELEFAX: 612-339-3061  
 ; INFORMATION FOR SEQ ID NO: 2:  
 ; SEQUENCE CHARACTERISTICS:  
 ; LENGTH: 294 amino acids  
 ; TYPE: amino acid  
 ; TOPOLOGY: linear  
 ; MOLECULE TYPE: protein  
 US-08-437-795-2

```

QY 211 LEKSVDFIASCISCGDEWNPQAEFTNAASTNSNLLLSYBEMLEKPVNTVKLAETWGC 270
      _ : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 157 F---VEKFMQGVYG-SWKYHKVSWWEKGSFVLFTFYEDKEDIKEVIKLIHFL-- 210
      _ : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 271 GFTDDEBKQGVDEIVKLCSDFNKQOVKNKGSVNSKIDNKH--FFRGEVYRDWANY 327
      _ : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 211 ---EKSEELVDRIIHHTSFQEKKNP-STNYTTLPDEIMNOKLSFPRKGGICDKNKH 266
      _ : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
QY 328 LTSEMIKLE 337
      _ : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db 267 FTVALNEXFD 276
      _ : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :

```

```

RESULT 8
US-09-609-816-10
  Sequence 10, Application US/09609816
  Patent No. 6436684
  GENERAL INFORMATION:
    APPLICANT: Woodage, Trevor
    APPLICANT: Wei, Minh Hui
    APPLICANT: Kodira, Chinappa
    APPLICANT: Beasley, Ellen
    APPLICANT: DiFrancesco, Valentina
    TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
    TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
    TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
    FILE REFERENCE: CLO00669PCT
    CURRENT APPLICATION NUMBER: US/09/609,816
    CURRENT FILING DATE: 2000-07-03
    PRIOR APPLICATION NUMBER: 60/192,408
    PRIOR FILING DATE: 2000-03-27
    PRIOR APPLICATION NUMBER: 60/212,725
    PRIOR FILING DATE: 2000-06-20
    PRIOR APPLICATION NUMBER: 09/609,816
    PRIOR FILING DATE: 2000-07-03
    NUMBER OF SEQ ID NOS: 23
    SOFTWARE: FaastSEQ for Windows Version 4.0
    SEQ ID NO 10
    LENGTH: 283
    TYPE: PRT
    ORGANISM: HUMAN
US-09-609-816-10

```

RESULT 9  
US-09-150-133-13  
; Sequence 13, Application US/09150133B  
; Patent No. 6060295

```

; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.504
; CURRENT APPLICATION NUMBER: US/09/150.133B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
US-09-150-133-13

Query Match 11.9%; Score 245.5; DB 3; Length 295;
Best Local Similarity 25.5%; Pred. No. 4.5e-15;
Matches 82; Conservative 48; Mismatches 131; Indels 61; Gaps 12;

QY 31 EQEKEEDSKMYKRYEIVSSLPNDYWGDTMRLYKGFQWQGYLPGIMAFEDNFKARETD 90
DB 2 ETSMPYEVVEFGFGRGLMDKRFKTKYWEDV-----EMFLARPDD 40

QY 91 ILITLTPKAGTTWTKALTFAILTR-DVNHPSPTPLFFNPHSCVQNLEYLYMGRENTM 149
DB 41 LVIAIYPSKGTWISSEVVMYIYKGDVEKCKEDA-----IFN-----RIPYLECRNEDLI 90

QY 150 PDLML--NESPLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI---- 203
DB 91 NGIKQLKEKESPRIVKTHLPKVLPAFWKCKMYLCRNKADKAVSYFYFLMLITSYP 150

QY 204 NPKKLDLEKSYDIPASGISFCGPEWPFQABFTNAASTNSNLLLSYEEMLEKPEVNVKK 263
DB 151 NPKSFSEF---VEKFMQGVPG-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVKK 206

QY 264 LAEFMCGGFTDDEEKQIGVDEIVKLCSPDNLKQOVNKGSSY-----NSKIDNKH 315
DB 207 LIEFL-----ERKPSAELVDRIIQTSTFQEMKNP-----STNYTMMPEEMNQKV--SPF 255

QY 316 FRKGEVRDWNALYLTSEMINKLE 337
DB 256 MRKGIIGDWKNHFFPEALRERFD 277

RESULT 10
US-09-150-141-13
; Sequence 13, Application US/09150141B
; Patent No. 6071732
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
; FILE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.495
; CURRENT APPLICATION NUMBER: US/09/150.141B
; CURRENT FILING DATE: 1998-09-09
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 5.1 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
US-09-150-141-13

Query Match 11.9%; Score 245.5; DB 3; Length 295;
Best Local Similarity 25.5%; Pred. No. 4.5e-15;
Matches 82; Conservative 48; Mismatches 131; Indels 61; Gaps 12;

QY 31 EQEKEEDSKMYKRYEIVSSLPNDYWGDTMRLYKGFQWQGYLPGIMAFEDNFKARETD 90
DB 2 ETSMPYEVVEFGFGRGLMDKRFKTKYWEDV-----EMFLARPDD 40

QY 91 ILITLTPKAGTTWTKALTFAILTR-DVNHPSPTPLFFNPHSCVQNLEYLYMGRENTM 149
DB 41 LVIAIYPSKGTWISSEVVMYIYKGDVEKCKEDA-----IFN-----RIPYLECRNEDLI 90

QY 150 PDLML--NESPLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI---- 203
DB 91 NGIKQLKEKESPRIVKTHLPKVLPAFWKCKMYLCRNKADKAVSYFYFLMLITSYP 150

QY 204 NPKKLDLEKSYDIPASGISFCGPEWPFQABFTNAASTNSNLLLSYEEMLEKPEVNVKK 263
DB 151 NPKSFSEF---VEKFMQGVPG-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVKK 206

QY 264 LAEFMCGGFTDDEEKQIGVDEIVKLCSPDNLKQOVNKGSSY-----NSKIDNKH 315
DB 207 LIEFL-----ERKPSAELVDRIIQTSTFQEMKNP-----STNYTMMPEEMNQKV--SPF 255

QY 316 FRKGEVRDWNALYLTSEMINKLE 337
DB 256 MRKGIIGDWKNHFFPEALRERFD 277

RESULT 11
US-09-374-493-13
; Sequence 13, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374.493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150.133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072.994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
US-09-374-493-13

Query Match 11.9%; Score 245.5; DB 3; Length 295;
Best Local Similarity 25.5%; Pred. No. 4.5e-15;
Matches 82; Conservative 48; Mismatches 131; Indels 61; Gaps 12;

QY 31 EQEKEEDSKMYKRYEIVSSLPNDYWGDTMRLYKGFQWQGYLPGIMAFEDNFKARETD 90
DB 2 ETSMPYEVVEFGFGRGLMDKRFKTKYWEDV-----EMFLARPDD 40

QY 91 ILITLTPKAGTTWTKALTFAILTR-DVNHPSPTPLFFNPHSCVQNLEYLYMGRENTM 149
DB 41 LVIAIYPSKGTWISSEVVMYIYKGDVEKCKEDA-----IFN-----RIPYLECRNEDLI 90

QY 150 PDLML--NESPLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI---- 203
DB 91 NGIKQLKEKESPRIVKTHLPKVLPAFWKCKMYLCRNKADKAVSYFYFLMLITSYP 150

QY 204 NPKKLDLEKSYDIPASGISFCGPEWPFQABFTNAASTNSNLLLSYEEMLEKPEVNVKK 263
DB 151 NPKSFSEF---VEKFMQGVPG-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVKK 206

QY 264 LAEFMCGGFTDDEEKQIGVDEIVKLCSPDNLKQOVNKGSSY-----NSKIDNKH 315
DB 207 LIEFL-----ERKPSAELVDRIIQTSTFQEMKNP-----STNYTMMPEEMNQKV--SPF 255
```

```

DB 2 ETSMPYEVVEFGFGRGLMDKRFKTKYWEDV-----EMFLARPDD 40
QY 91 ILITLTPKAGTTWTKALTFAILTR-DVNHPSPTPLFFNPHSCVQNLEYLYMGRENTM 149
DB 41 LVIAIYPSKGTWISSEVVMYIYKGDVEKCKEDA-----IFN-----RIPYLECRNEDLI 90
QY 150 PDLML--NESPLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI---- 203
DB 91 NGIKQLKEKESPRIVKTHLPKVLPAFWKCKMYLCRNKADKAVSYFYFLMLITSYP 150
QY 204 NPKKLDLEKSYDIPASGISFCGPEWPFQABFTNAASTNSNLLLSYEEMLEKPEVNVKK 263
DB 151 NPKSFSEF---VEKFMQGVPG-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVKK 206
QY 264 LAEFMCGGFTDDEEKQIGVDEIVKLCSPDNLKQOVNKGSSY-----NSKIDNKH 315
DB 207 LIEFL-----ERKPSAELVDRIIQTSTFQEMKNP-----STNYTMMPEEMNQKV--SPF 255

RESULT 11
US-09-374-493-13
; Sequence 13, Application US/09374493
; Patent No. 6204016
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.546
; CURRENT APPLICATION NUMBER: US/09/374.493
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150.133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072.994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
US-09-374-493-13

Query Match 11.9%; Score 245.5; DB 3; Length 295;
Best Local Similarity 25.5%; Pred. No. 4.5e-15;
Matches 82; Conservative 48; Mismatches 131; Indels 61; Gaps 12;

QY 31 EQEKEEDSKMYKRYEIVSSLPNDYWGDTMRLYKGFQWQGYLPGIMAFEDNFKARETD 90
DB 2 ETSMPYEVVEFGFGRGLMDKRFKTKYWEDV-----EMFLARPDD 40

QY 91 ILITLTPKAGTTWTKALTFAILTR-DVNHPSPTPLFFNPHSCVQNLEYLYMGRENTM 149
DB 41 LVIAIYPSKGTWISSEVVMYIYKGDVEKCKEDA-----IFN-----RIPYLECRNEDLI 90

QY 150 PDLML--NESPLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI---- 203
DB 91 NGIKQLKEKESPRIVKTHLPKVLPAFWKCKMYLCRNKADKAVSYFYFLMLITSYP 150

QY 204 NPKKLDLEKSYDIPASGISFCGPEWPFQABFTNAASTNSNLLLSYEEMLEKPEVNVKK 263
DB 151 NPKSFSEF---VEKFMQGVPG-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVKK 206
QY 264 LAEFMCGGFTDDEEKQIGVDEIVKLCSPDNLKQOVNKGSSY-----NSKIDNKH 315
DB 207 LIEFL-----ERKPSAELVDRIIQTSTFQEMKNP-----STNYTMMPEEMNQKV--SPF 255
```

```

QY 316 FRKGEVRDWANVLTSEMICKLE 337
Db 256 MRKGIIGDWKNHFFPEALRERFD 277

RESULT 12
US-09-374-824-13
; Sequence 13, Application US/09374824
; Patent No. 6207414
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.547
; CURRENT APPLICATION NUMBER: US/09/374,824
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,133
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
; US-09-374-824-13

Query Match 11.9%; Score 245.5; DB 3; Length 295;
Best Local Similarity 25.5%; Pred. No. 4.5e-15;
Matches 82; Conservative 48; Mismatches 131; Indels 61; Gaps 12;

QY 31 EQEKEEDSKMYKRYREIVTSSLPNDYMGDTMRLYKGFWMQGYLVPGIMAFEDNFKARETD 90
Db 2 ETSMPYEYVFGFGRGVLMDKRTKYWEDV-----ENFLARPDD 40
QY 91 IILTLPRKAGTWTKALTFAILTR-DVNHPSSTPHLLFPNPHSCVONLEYLYMGRENTM 149
Db 41 LVIAATPKSGTTWISVVVMYIKEGDVCKEDA-----IFN-----RIPYLECRNEDLI 90
QY 150 PDLMDL--NESPRLFAGHIPVSLPASVLKSGTKIINISNRKSTFVSFWKFGNLI----- 203
Db 91 NGIKQLKEKESPRIVKTHLPKVLPAFWKNCCKMIVLCRNADKDVAVSYFFLLMITSYP 150
QY 204 NPKSFSBP---VEKFMQGVYPG-SWYDHVKAWKSKNSRVLFMFYEDMKEDIRREVVK 263
Db 151 NPKSFSBP---VEKFMQGVYPG-SWYDHVKAWKSKNSRVLFMFYEDMKEDIRREVVK 206
QY 264 LAEPWCGGFTDDEKQGVDEIVKLCSPDNLKNOQVKNKGSY-----NSKIDNKH 315
Db 207 LIEFL-----BRKPSAELVDRIIHTSFQEMKNRP-----STNYTMPEEMMNQKV--SPF 255
QY 316 FRKGEVRDWANVLTSEMICKLE 337
Db 256 MRKGIIGDWKNHFFPEALRERFD 277

RESULT 13
US-09-374-492-13
; Sequence 13, Application US/09374492
; Patent No. 6207432
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5820.545
; CURRENT APPLICATION NUMBER: US/09/374,492
; CURRENT FILING DATE: 1999-08-13
; EARLIER APPLICATION NUMBER: 09/150,141
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994
; EARLIER FILING DATE: 1998-09-09
; EARLIER APPLICATION NUMBER: 60/072,994

```

```

; EARLIER FILING DATE: 1998-01-29
; EARLIER APPLICATION NUMBER: PCT/US99/16750
; EARLIER FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
; US-09-374-492-13

Query Match 11.9%; Score 245.5; DB 3; Length 295;
Best Local Similarity 25.5%; Pred. No. 4.5e-15;
Matches 82; Conservative 48; Mismatches 131; Indels 61; Gaps 12;

QY 31 EQEKEEDSKMYKRYREIVTSSLPNDYMGDTMRLYKGFWMQGYLVPGIMAFEDNFKARETD 90
Db 2 ETSMPYEYVFGFGRGVLMDKRTKYWEDV-----ENFLARPDD 40
QY 91 IILTLPRKAGTWTKALTFAILTR-DVNHPSSTPHLLFPNPHSCVONLEYLYMGRENTM 149
Db 41 LVIAATPKSGTTWISVVVMYIKEGDVCKEDA-----IFN-----RIPYLECRNEDLI 90
QY 150 PDLMDL--NESPRLFAGHIPVSLPASVLKSGTKIINISNRKSTFVSFWKFGNLI----- 203
Db 91 NGIKQLKEKESPRIVKTHLPKVLPAFWKNCCKMIVLCRNADKDVAVSYFFLLMITSYP 150
QY 204 NPKSFSBP---VEKFMQGVYPG-SWYDHVKAWKSKNSRVLFMFYEDMKEDIRREVVK 263
Db 151 NPKSFSBP---VEKFMQGVYPG-SWYDHVKAWKSKNSRVLFMFYEDMKEDIRREVVK 206
QY 264 LAEPWCGGFTDDEKQGVDEIVKLCSPDNLKNOQVKNKGSY-----NSKIDNKH 315
Db 207 LIEFL-----BRKPSAELVDRIIHTSFQEMKNRP-----STNYTMPEEMMNQKV--SPF 255
QY 316 FRKGEVRDWANVLTSEMICKLE 337
Db 256 MRKGIIGDWKNHFFPEALRERFD 277

RESULT 14
US-09-785-343-13
; Sequence 13, Application US/09785343
; Patent No. 6605455
; GENERAL INFORMATION:
; APPLICANT: The Board of Regents of the University of Oklahoma
; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES AND METHODS OF USE THEREOF
; FILE REFERENCE: 5827.003
; CURRENT APPLICATION NUMBER: US/09/785,343
; CURRENT FILING DATE: 2001-02-16
; PRIOR APPLICATION NUMBER: 09/150,133
; PRIOR FILING DATE: 1998-09-09
; PRIOR APPLICATION NUMBER: 60/072,994
; PRIOR FILING DATE: 1998-01-29
; PRIOR APPLICATION NUMBER: PCT/US99/16750
; PRIOR FILING DATE: 1999-07-23
; NUMBER OF SEQ ID NOS: 45
; SOFTWARE: Wordperfect 8.0 (saved in ASCII format)
; SEQ ID NO 13
; LENGTH: 295
; TYPE: PRT
; ORGANISM: Mus musculus
; FEATURE:
; OTHER INFORMATION: estrogen sulfotransferase
; US-09-785-343-13

Query Match 11.9%; Score 245.5; DB 4; Length 295;
Best Local Similarity 25.5%; Pred. No. 4.5e-15;
Matches 82; Conservative 48; Mismatches 131; Indels 61; Gaps 12;

QY 31 EQEKEEDSKMYKRYREIVTSSLPNDYMGDTMRLYKGFWMQGYLVPGIMAFEDNFKARETD 90

```

Db 2 ETSMEPEYVEGEFGRVLMKRFKTYWEDV-----EMFLARPDD 40  
QY 91 IILITLPRAGTTWTKALFAILLTR-DVNHPSPTPLFFNPHSCVQNLLEYLYMGRENTM 149  
Db 41 LVIAIYPSKGTIWIWSEVIMYIKGDUKCKEDA-----IFN-----RIPYLECRNEDLI 90  
QY 150 PDLMDL--NESPLPAGHIPYSLLPASVLKSGTKIINISNRKSTFTVSWFKGNLI----- 203  
Db 91 NGIKQLKEKESPRIVKTHLPKVLPAFWEKNCCKMILCRNAKOVAVSYYYFLMLITSYP 150  
QY 204 NPKKLLDLEKSDVDFASGISFCGPENNFQAEFTNAASTNSNLLLSYEEMLEKPVENVYK 263  
Db 151 NPKSPSEF---VEKMQGQVPG-SWYDHVKAWEKSKNSRVLFMFYEDMKEDIRREVVK 206  
QY 264 LAEFMCGCFDDEEKQIGVDEIVKLCSFDNLKQVKNKNGSSY-----NSKIDNKH 315  
Db 207 LIEFL-----ERKPSAELVDRIIQTSTFQEMCNPF-----STNYTWPPEMMNQV--SPF 255  
QY 316 FRKGEVRDWANYLTSEMIXKLE 337  
Db 256 MRKGIIGDWKNHFPALRERFD 277

RESULT 15

US-09-795-926-6  
; Sequence 6, Application US/09795926  
; Patent No. 6555669  
; GENERAL INFORMATION:  
; APPLICANT: Donoho, Gregory  
; APPLICANT: Hilbun, Erin  
; APPLICANT: Turner, C. Alexander Jr.  
; APPLICANT: Friedrich, Glenn  
; APPLICANT: Abuin, Alejandro  
; APPLICANT: Zambrowicz, Brian  
; APPLICANT: Sands, Arthur T.  
; APPLICANT: Walke, D. Wade  
; APPLICANT: Wilganowski, Nathaniel L.  
; APPLICANT: Hu, Yi  
; APPLICANT: Kieke, James Alvin  
; APPLICANT: Potter, David George  
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND  
; FILE REFERENCE: LEX-0144-USA  
; CURRENT APPLICATION NUMBER: US/09/795,926  
; CURRENT FILING DATE: 2001-02-28  
; PRIOR APPLICATION NUMBER: US 60/185,920  
; PRIOR FILING DATE: 2000-02-29  
; PRIOR APPLICATION NUMBER: US 60/186,558  
; PRIOR FILING DATE: 2000-03-02  
; PRIOR APPLICATION NUMBER: US 60/191,849  
; PRIOR FILING DATE: 2000-03-24  
; NUMBER OF SEQ ID NOS: 47  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 265  
; TYPE: PRT  
; ORGANISM: homo sapiens  
US-09-795-926-6

Query Match 11.6%; Score 239; DB 4; Length 265;  
Best Local Similarity 27.1%; Pred No. 1.6e-14;  
Matches 79; Conservative 46; Mismatches 121; Indels 46; Gaps 12;  
QY 82 DNFKARETDIILTLPRAGTTW---TKALTFAILTRDVNHPSPTPLFFNPHSCVQN 137  
Db 11 DTFEARHDDIVLASYPKCGSNWILHIVSELIYAVSKKKYKYPFPV-----LECGDS 62  
QY 138 LEVLYMGRENTMPDLMLNESPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFTVSW 197  
Db 63 EKY---QRMKGFP-----SPRILATHLYDKLPGSIFENKAKILVIFRNPKDTAVSFL 112  
QY 198 KFGNLIINPKLLDLEKSDVDFASGISFCGPENNFQAEFT---NAASTNSNLLLSYEEM 254

Db 113 HFHNDV-PD--IPSYGSWDEFFRQPMKQVSGRYPDFDFAINWKNHLDGDNVFKFELYEDLK 169  
QY 255 EKFEVENVKKLAEFMCGCFDDEEKQIGVDEIVKLCSFDNLKQVKNKNGSSYNSKIDNKH 314  
Db 170 ENLAAGIKQIAEFLGPFLLTGEQ-----IQTISVQSTFQAMRAKSQDTHGAV-----GPF 218  
QY 315 FRKGEVRDWANYLTSEMIXKLETAGKINSEKHLISNKNL-TNFNLNNYCE 365  
Db 219 LFRKGEVGDWKN-LFSEIQNQ-----EMDEKFKECUAGTSLGAKLKYESYCO 264

Search completed: March 20, 2004, 01:39:19  
Job time : 18 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:37:48 ; Search time 34 Seconds  
(without alignments)  
2977.986 Million cell updates/sec

Title: US-09-854-122-16  
Perfect score: 2059  
Sequence: 1 TRGTGIAVACLPLIMAGIL.....CLLSYSFVINNFILLKXXXX 391

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1049977 seqs, 258955339 residues

Total number of hits satisfying chosen parameters: 1049977

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US05\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US05\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep.\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2059	100.0	391	9	US-09-854-122-16
2	1750	85.0	331	9	US-09-854-122-17
3	589	28.6	344	12	US-10-424-599-201713
4	527.5	25.6	302	9	US-09-854-122-19
5	525.5	25.6	365	12	US-10-425-114-63828
6	525.5	25.3	365	12	US-10-425-114-65063
7	521.5	25.3	338	15	US-10-259-194A-4
8	519.5	25.2	348	15	US-10-259-194A-380
9	510.5	24.8	320	9	US-09-854-122-20
10	501.5	24.4	328	14	US-10-259-165-114
11	501.5	24.4	328	14	US-10-259-165-444
12	500.5	24.3	324	9	US-09-854-122-18
13	483	23.5	688	14	US-10-195-144-79
14	483	23.5	688	15	US-10-345-072-79
15	479	23.3	343	15	US-10-259-194A-330

16	446.5	21.7	346	14	US-10-259-165-332	Sequence 332, App
17	444.5	21.6	251	12	US-10-424-599-225084	Sequence 225084,
18	441.5	21.4	486	15	US-10-259-194A-132	Sequence 132, App
19	363.5	17.7	349	15	US-10-259-194A-16	Sequence 16, Appl
20	309	15.0	269	12	US-10-072-012-855	Sequence 855, App
21	309	15.0	269	12	US-10-072-012-857	Sequence 867, App
22	304	14.8	302	12	US-10-072-012-586	Sequence 586, App
23	288	13.0	304	13	US-10-193-330-6	Sequence 6, Appl
24	288	13.0	304	14	US-10-193-334-6	Sequence 6, Appl
25	268	13.0	304	14	US-10-193-329-8	Sequence 585, App
26	266.5	12.9	307	12	US-10-072-012-585	Sequence 206, App
27	265.5	12.9	295	12	US-10-072-012-206	Sequence 582, App
28	264	12.8	304	12	US-10-072-012-582	Sequence 4, Appl
29	264	12.8	304	13	US-10-193-330-4	Sequence 4, Appl
30	264	12.8	304	14	US-10-193-334-4	Sequence 4, Appl
31	264	12.8	304	14	US-10-193-329-4	Sequence 5, Appl
32	260	12.6	304	13	US-10-193-330-5	Sequence 5, Appl
33	260	12.6	304	14	US-10-193-334-5	Sequence 5, Appl
34	260	12.6	304	14	US-10-193-329-5	Sequence 5, Appl
35	255.5	12.4	283	9	US-09-898-570-40	Sequence 40, Appl
36	255.5	12.4	283	10	US-09-839-446-40	Sequence 40, Appl
37	255.5	12.4	283	13	US-10-193-330-8	Sequence 8, Appl
38	255.5	12.4	283	13	US-10-193-330-9	Sequence 8, Appl
39	255.5	12.4	283	14	US-10-193-334-8	Sequence 8, Appl
40	255.5	12.4	283	14	US-10-193-334-9	Sequence 8, Appl
41	255.5	12.4	283	14	US-10-193-329-9	Sequence 8, Appl
42	255.5	12.4	283	14	US-10-193-329-9	Sequence 9, Appl
43	255.5	12.4	304	12	US-10-072-012-583	Sequence 583, App
44	253.5	12.3	305	9	US-09-898-570-26	Sequence 26, Appl
45	253.5	12.3	305	10	US-09-839-446-26	Sequence 26, Appl

ALIGNMENTS

RESULT 1

US-09-854-122-16  
; Sequence 16, Application US/09854122  
; Patent No. US20020016980A1  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, ROBERT  
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
; FILE REFERENCE: PRA-007.01  
; CURRENT APPLICATION NUMBER: US/09/854,122  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/202,529  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 16  
; LENGTH: 391  
; TYPE: PRT  
; ORGANISM: Zostera marina  
; US-09-854-122-16

Query Match	100.0%	Score	2059	DB 9	Length	391	
Best Local Similarity	100.0%	Pred. No.	4.5e-185				
Matches	391	Conservative	0	Mismatches	0	Gaps	0
Qy	1	TRGTGIAVACLPLIMAGILALEKFCGSKNEQKEEDSKMYKYRIRIVSSLPNDYWGDT	60				
Db	1	TRGTGIAVACLPLIMAGILALEKFCGSKNEQKEEDSKMYKYRIRIVSSLPNDYWGDT	60				
Qy	61	MLRYKGFQWQGYLVPGIMAFEDNFKARETDIILLTTPKAGTTWTKALTFAILLTRDVNHP	120				
Db	61	MLRYKGFQWQGYLVPGIMAFEDNFKARETDIILLTTPKAGTTWTKALTFAILLTRDVNHP	120				
Qy	121	SPTHPLFFNFHSCVQNLVYMGRENTMPDMLNESPRLFAGHIPYSLLPASVLKSGT	180				
Db	121	SPTHPLFFNFHSCVQNLVYMGRENTMPDMLNESPRLFAGHIPYSLLPASVLKSGT	180				
Qy	181	KIINISRNKSTFTSVFKWFGNLPDMLNESPRLFAGSIFSGCPGPNWQAEFTNAAS	240				

Db 181 KIINSTRKSKTSFVFWKPGNINPKLLDLEKSDVDFASGISFCGPEWFOAFTNAAS 240  
QY 241 TNSNLLLSYEEMLEKPVENVKLAEFMGCGFTDDEKQGVDEIVKLCSPDNLKQOVN 300  
Db 241 TNSNLLLSYEEMLEKPVENVKLAEFMGCGFTDDEKQGVDEIVKLCSPDNLKQOVN 300  
QY 301 KNGSSVNSKIDNKHFRKGEVRDWNALYLTSEMIKKTETAGKINSEKHLNKNLTNFKL 360  
Db 301 KNGSSVNSKIDNKHFRKGEVRDWNALYLTSEMIKKTETAGKINSEKHLNKNLTNFKL 360  
QY 361 NNYCKSNISLCLSYFVNNFILLKKKK 391  
Db 361 NNYCKSNISLCLSYFVNNFILLKKKK 391

RESULT 2  
US-09-854-122-17  
; Sequence 17, Application US/09854122  
; Patent No. US20020016980A1  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, ROBERT  
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
; FILE REFERENCE: PHA-007.01  
; CURRENT APPLICATION NUMBER: US/09/854,122  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 17  
; LENGTH: 331  
; TYPE: PRT  
; ORGANISM: Zostera marina  
US-09-854-122-17

Query Match 85.0%; Score 1750; DB 9; Length 331;  
Best Local Similarity 100.0%; Pred. No. 4.4e-156;  
Matches 331; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 16 MAGILALEKCFGSKNEQKEEDSKMYKRYRIVSLPSNDYWGDTMRLYKGFQWQGYLVP 75  
Db 1 MAGILALEKCFGSKNEQKEEDSKMYKRYRIVSLPSNDYWGDTMRLYKGFQWQGYLVP 60

QY 76 GIMAFEDNFKARETDIITLTPKAGTTWKALTFAITRDVNHPSPTPLFFNPHSCV 135  
Db 61 GIMAFEDNFKARETDIITLTPKAGTTWKALTFAITRDVNHPSPTPLFFNPHSCV 120

QY 136 ONLEYLYMGRENTMPDMLNESPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFVS 195  
Db 121 ONLEYLYMGRENTMPDMLNESPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFVS 180

QY 196 FWKFGNLIINPKLLDLEKSDVDFASGISFCGPEWFOAFTNAASTNSNLLLSYEEMLE 255  
Db 181 FWKFGNLIINPKLLDLEKSDVDFASGISFCGPEWFOAFTNAASTNSNLLLSYEEMLE 240

QY 256 KPVENVKLAEFMGCGFTDDEKQGVDEIVKLCSPDNLKQOVNKGSSVNSKIDNKH 315  
Db 241 KPVENVKLAEFMGCGFTDDEKQGVDEIVKLCSPDNLKQOVNKGSSVNSKIDNKH 300

QY 316 FRKGEVRDWNALYLTSEMIKKTETAGKINSE 346  
Db 301 FRKGEVRDWNALYLTSEMIKKTETAGKINSE 331

RESULT 3  
US-10-424-599-201713  
; Sequence 201713, Application US/10424599  
; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 201713  
; LENGTH: 344  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_24172C.1.pap  
US-10-424-599-201713

Query Match 28.6%; Score 589; DB 12; Length 344;  
Best Local Similarity 38.6%; Pred. No. 1e-46;  
Matches 125; Conservative 69; Mismatches 116; Indels 14; Gaps 8;

QY 25 CFGSKNEQKE-----EDSKMYKRYRIVSLPSNDYWGDT-MRLYKGFQWQGYLVP 78  
Db 8 CPREENSEKGEITIEDKLSQCKELILSLPRGWRTRYLYLFGQFWCQPLEIQAI 67

QY 79 AFEDNFKARETDIITLTPKAGTTWKALTFAITRDVNH--PSSPTPLFFNPHSCV 136  
Db 68 TFQHFQAKSDVIVATIPKSGITWLTALFAIVNRHTSITSSMSHPLTSPHELVP 127

QY 137 NLEY-LYMGRENTMPDMLNESPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFVS 195  
Db 128 FIEVTYGNAPSHVPLNSNMT-PRLFQTHIPHALAKSIKESNRIIVICRNPLDTFVS 186

QY 196 FWKFGNLIINPKLLDLEKSDVDFASGISFCGPEWFOAFTNAASTNSNLLLSYE 252  
Db 187 TWIFLNKIPKLPPELGEAFKPKYKGIIGFTWQMLGYWKESIAIARPSKVLFLK 246

QY 253 MLEKPVENVKLAEFMGCGFTDDEKQGVDEIVKLCSPDNLKQOVNKGSSVNSKIDN 312  
Db 247 LKQVNVFHVKIAEFGLGPTSEBEGDTIESIILCSPEKKELEANKSG-TFARNFER 305

QY 313 KHFRKGEVRDWNALYLTSEMIKKL 336  
Db 306 KILFRKAEWGDVWNYLSPFEMGEKL 329

RESULT 4  
US-09-854-122-19  
; Sequence 19, Application US/09854122  
; Patent No. US20020016980A1  
; GENERAL INFORMATION:  
; APPLICANT: ALBERT, RANDALL S.  
; APPLICANT: SMITH, ROBERT  
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
; FILE REFERENCE: PHA-007.01  
; CURRENT APPLICATION NUMBER: US/09/854,122  
; CURRENT FILING DATE: 2001-09-10  
; PRIOR APPLICATION NUMBER: 60/202,529  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ ID NOS: 51  
; SOFTWARE: Patent In Ver. 2.1  
; SEQ ID NO 19  
; LENGTH: 302  
; TYPE: PRT  
; ORGANISM: Arabidopsis thaliana  
US-09-854-122-19

Query Match 25.6%; Score 527.5; DB 9; Length 302;  
Best Local Similarity 41.1%; Pred. No. 5.3e-41;  
Matches 122; Conservative 47; Mismatches 107; Indels 21; Gaps 9;

QY 37 DSKMYKRYRIVSLPSNDYWGDTMRLYKGFQWQGYLVPFIMAFEDNFKARETDIILT 95  
Db 13 DEDLTQETRALISSLPKEKGLWVSEIVFQGLWHTQAILQILQCKRFEAKDSIILVT 72

```
QY 96 LPKAGTTWKALTFAILTRDVNH-----PSSPTPLLPFNPHSCVONLEYLYMGRENTMPD 151
DB 73 NPKSGTTWKALVFALLNR---HKPPVSSSGNHPLLVNTHPLLVPFLEGVYV-----ESPD 125
QY 152 LDMLN-ESPRLPAGHIFVSLPASVILKSGTKIINISNRKSTFVSWFKGNLNPDKLLD 210
DB 126 FDFSLPSPRLNTHISLSLPESVSSCKIVCCRNPKDMFVSLWHFGKLAPEETAD 185
QY 211 --LEKSVDFIFASGIFCGPEWNFQAEFTNAASTNSN-LLLSYEEMLEKPVENYKLAEF 267
DB 186 YPIEKAVEAFCEGKFIGGFWDHILEVYASRENPKVLFVYTBELKKQTEVEMKRIAEF 245
QY 268 MCGGTDEEEKQIVDEIVKLCSPDNLKQVKNKNGSSYNKIDNKHFFRKGEVRDW 324
DB 246 LECGFIEBEE---VREIVKLCSPESLSNLEVNKGKLPNG-IETKFFRKGEIGGW 297

RESULT 5
US-10-425-114-63828
; Sequence 63828, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63828
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3353-009-H8_FLI.pep
US-10-425-114-63828
Query Match 25.5%; Score 525.5; DB 12; Length 365;
Best Local Similarity 35.2%; Pred. No. 1.1e-40;
Matches 116; Conservative 65; Mismatches 114; Indels 35; Gaps 11;
QY 30 NQOEKEED--SKMYKRYRIVSGLPSNDYWGDT-----MRLYKGFWMQGYL----VPGI 77
DB 31 NQAEASPLASPHSNIAIISLPL-----ETRCPPFLRYANFWVPEVILKADLPGI 85
QY 78 MAFEDNFKARETDIILTLPKAGTTWKALTFAILTRDVNHSPSSPTPLLPFNPHSCVON 137
DB 86 ---HSCFKPRPTDVFVASFPKSGTTWLKALAFATLKRSTHPLDGDHPLRRCNPHDCVRF 142
QY 138 LEVLYMGRENTMPDLDMLNESPRLPAGHIFVSLPASVL-----KSGTKIINISNRKSTF 193
DB 143 LDANFNQCKDEAL-----PSRVLATHLPYLLPGSITGDRERSGCRIVYVCRPKDAL 198
QY 194 VSPWKF----GNLINPD-KLLDLEKSVDFIFASGIFCGPEWNFQAEF-TNAASTNSNLL 247
DB 199 VSAWLFRKAASALGADARSFTIQEALFCDGRCMCGPQWEHVLYQWBEESVRRPDRVLF 258
QY 248 LSYEEMLEKPVENYKLAEFMCGGFTDDEEKQIVDEIVKLCSPDNLKQVKNKNGSSYN 307
DB 259 LRYEEMLIDPEAHVRKLAIFMCGGFBEBEHEHGVSAIVELCSLGNKRDMEVNRNGSNML 318
QY 308 SKIDNKHFFRKGEVRDWANYLTSEMICKLE 337
DB 319 G-VKNESYFRKGVAGDSNEMTDMQRLD 347

RESULT 6
US-10-425-114-63828
; Sequence 63828, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 63828
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB3353-009-H8_FLI.pep
US-10-425-114-63828
Query Match 25.5%; Score 525.5; DB 12; Length 365;
Best Local Similarity 35.2%; Pred. No. 1.1e-40;
Matches 116; Conservative 65; Mismatches 114; Indels 35; Gaps 11;
QY 30 NQOEKEED--SKMYKRYRIVSGLPSNDYWGDT-----MRLYKGFWMQGYL----VPGI 77
DB 31 NQAEASPLASPHSNIAIISLPL-----ETRCPPFLRYANFWVPEVILKADLPGI 85
QY 78 MAFEDNFKARETDIILTLPKAGTTWKALTFAILTRDVNHSPSSPTPLLPFNPHSCVON 137
DB 86 ---HSCFKPRPTDVFVASFPKSGTTWLKALAFATLKRSTHPLDGDHPLRRCNPHDCVRF 142
QY 138 LEVLYMGRENTMPDLDMLNESPRLPAGHIFVSLPASVL-----KSGTKIINISNRKSTF 193
DB 143 LDANFNQCKDEAL-----PSRVLATHLPYLLPGSITGDRERSGCRIVYVCRPKDAL 198
QY 194 VSPWKF----GNLINPD-KLLDLEKSVDFIFASGIFCGPEWNFQAEF-TNAASTNSNLL 247
DB 199 VSAWLFRKAASALGADARSFTIQEALFCDGRCMCGPQWEHVLYQWBEESVRRPDRVLF 258
QY 248 LSYEEMLEKPVENYKLAEFMCGGFTDDEEKQIVDEIVKLCSPDNLKQVKNKNGSSYN 307
DB 259 LRYEEMLIDPEAHVRKLAIFMCGGFBEBEHEHGVSAIVELCSLGNKRDMEVNRNGSNML 318
QY 308 SKIDNKHFFRKGEVRDWANYLTSEMICKLE 337
DB 319 G-VKNESYFRKGVAGDSNEMTDMQRLD 347
```

```
US-10-425-114-65063
; Sequence 65063, Application US/10425114
; Publication No. US20040034888A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE REFERENCE: 38-21(5313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 65063
; LENGTH: 365
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4740-009-D3_FLI.pep
US-10-425-114-65063
Query Match 25.5%; Score 525.5; DB 12; Length 365;
Best Local Similarity 35.2%; Pred. No. 1.1e-40;
Matches 116; Conservative 65; Mismatches 114; Indels 35; Gaps 11;
QY 30 NQOEKEED--SKMYKRYRIVSGLPSNDYWGDT-----MRLYKGFWMQGYL----VPGI 77
DB 31 NQAEASPLASPHSNIAIISLPL-----ETRCPPFLRYANFWVPEVILKADLPGI 85
QY 78 MAFEDNFKARETDIILTLPKAGTTWKALTFAILTRDVNHSPSSPTPLLPFNPHSCVON 137
DB 86 ---HSCFKPRPTDVFVASFPKSGTTWLKALAFATLKRSTHPLDGDHPLRRCNPHDCVRF 142
QY 138 LEVLYMGRENTMPDLDMLNESPRLPAGHIFVSLPASVL-----KSGTKIINISNRKSTF 193
DB 143 LDANFNQCKDEAL-----PSRVLATHLPYLLPGSITGDRERSGCRIVYVCRPKDAL 198
QY 194 VSPWKF----GNLINPD-KLLDLEKSVDFIFASGIFCGPEWNFQAEF-TNAASTNSNLL 247
DB 199 VSAWLFRKAASALGADARSFTIQEALFCDGRCMCGPQWEHVLYQWBEESVRRPDRVLF 258
QY 248 LSYEEMLEKPVENYKLAEFMCGGFTDDEEKQIVDEIVKLCSPDNLKQVKNKNGSSYN 307
DB 259 LRYEEMLIDPEAHVRKLAIFMCGGFBEBEHEHGVSAIVELCSLGNKRDMEVNRNGSNML 318
QY 308 SKIDNKHFFRKGEVRDWANYLTSEMICKLE 337
DB 319 G-VKNESYFRKGVAGDSNEMTDMQRLD 347

RESULT 7
US-10-259-194A-4
; Sequence 4, Application US/10259194A
; Publication No. US2004010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Grasseman, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
```

```
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 4
; LENGTH: 338
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-194A-4

Query Match      25.3%; Score 521.5; DB 15; Length 338;
Best Local Similarity 38.4%; Pred. No. 2.3e-40;
Matches 118; Conservative 54; Mismatches 116; Indels 19; Gaps 8;

QY 42 KRYREIVSSLPNDYWGDTMRLYKGFQWQGVLPVPGIMAFEDNF-KARETDIILTLTKAG 100
DB 33 EEVAGVSSLP--YPKLEHYQWMLMEYTLFGIMAIQSFVPRRHGDVVLASPGKCG 90
QY 101 TTWTKALTFAITRDVNHSSPTLFFNPHSCVQNVLEYL-----YMGRENTMPDLMLN 156
DB 91 TTWLKALAFVARGAYSPASDRHPLRLNPHDCCVPFMEGAISEGWGGKIDELP----- 144
QY 157 ESPRLPAGHIPYSLLPASVYL--KSGTKIINISNRKSTFVSPFKGNLINPDKLLDLEKSV 215
DB 145 -SPRLMSTWQHAALPKSIADSPGCKVYICREPKDILVSAWHFRRIEED--LSFGVEF 201
QY 216 DIPASGISCGEWFNFQAEFTNAASTN--SNLLLSYEEMLEKPVENVKLAEPFGCGFTD 274
DB 202 EAACDCKFTGAIWHIIGYWACKANPEKVLFLVYEDLLRDPANIVRKLADFLGQPFSS 261
QY 275 PREKQIVDEIVKLSFDNLKNQVKNKGSSVNSKIDNKHFRKGEVRDWNVLYTSEMIX 334
DB 262 TEERGLVTDIVRLCSFNKLSLVNKGEA--SPAPPNASYFRKGAGDKWHHTPEWVE 320
QY 335 KLETACK 341
DB 321 CFTDITVK 327
```

```
RESULT 8
US-10-259-194A-380
; Sequence 380, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghasseman, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; CURRENT FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
```

```
; SEQ ID NO 380
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Oryza sativa
US-10-259-194A-380

Query Match      25.2%; Score 519.5; DB 15; Length 348;
Best Local Similarity 40.3%; Pred. No. 3.7e-40;
Matches 120; Conservative 48; Mismatches 117; Indels 13; Gaps 8;

QY 48 VSSLPNDYWGDTMRLYKGFQWQGVLPVPGIMAF--EDNFKARETDIILTLTKAGTTWTKA 106
DB 44 VSAUPAGVSYGQPCRCYGGTWVFSWAQGVVAMHRRGLVPRAGDVLLASLPKSGTTWTKA 103
QY 107 LTFAILTRDV--NHPSSPTLFFNPHSCVQNVLEYL--MGRENTMPDLMLNESPRLPAG 164
DB 104 LAFATMARACRPPPPASDRHPLRLNPHDCCVFLDLRLFAAGRDALDEL-----ESPRLMCT 159
QY 165 HIPYSLLPASVYL--KSGTKIINISNRKSTFVSPFKGNLINPDKLLDLEKSVDFASGI 222
DB 160 HMPJLSLPPAVADGNSNTKIYICDQDKRLVSIWHPRKENVFDLL--LQEVYESICDGT 217
QY 223 SFCGPEWNFQAEFTNAASTN--SNLLLSYEEMLEKPVENVKLAEPFGCGFTDDEKQGI 281
DB 218 GPAGVMDHLLGYWRASKIDILGRVLFKYEVLDPVNTVRELAQFVGQPFSDTESEAGI 277
QY 282 VDEIVKLSFDNLKNQVKNKG--SSVNSKIDNKHFRKGEVRDWNVLYTSEMIXKET 338
DB 278 VAEIVKLSLESLSQKANKREGIOGVYIKESHDSYFRKGVGEGDWRNMTPKGHEHLS 335

RESULT 9
US-09-854-122-20
; Sequence 20, Application US/09854122
; Patent No. US20020016980A1
; GENERAL INFORMATION:
; APPLICANT: ALBERTE, RANDALL S.
; APPLICANT: SMITH, ROBERT
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA
; FILE REFERENCE: PHA-007.01
; CURRENT APPLICATION NUMBER: US/09/854,122
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/202,529
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Flaveria bidentis
US-09-854-122-20

Query Match      24.8%; Score 510.5; DB 9; Length 320;
Best Local Similarity 37.6%; Pred. No. 2.3e-39;
Matches 120; Conservative 58; Mismatches 110; Indels 31; Gaps 9;

QY 35 EEDSKYKRYEIVSSLP----SNDYWGDTMRLYKGFQWQGVLPVPGIMAFEDNFKARETD 90
DB 2 ETTTKQFESMAEMIKCLPQHTCSSLKGRTILYKQDFWGLQNNIEGAILAQQSFKARPD 61
QY 91 IILTLTKAGTTWTKALTFAITRDVNHSPSPHLLFFNPHSCVQNVLEYLYMGRENTMP 150
DB 62 VFLCSYPSAGTTWALKAYALVIRE--KDFEFTSPLLTNIPNCIPIE-----K 109
QY 151 LDML--NESRPLF---AGHIPYSLLPASVLSKGTGKIINISNRKSTFTFSFWFGNLMP 205
DB 110 DLKKIVENQNSCFTPMATHMPYHVLPSKILANCKMVIYIRNKIVIVSYFHFGREITK 169
QY 206 DKLLD--LEKSVDFISGIFCGPEWNFQAEFTNAASTNSNLL--LSYEEMLEKPVENVK 262
DB 170 LPLEDAPEEAFDFYHGISQFGPYWDHLLGYWKASLEREVEVILFKYEDVKDPSNVK 229
QY 263 KLAEPFGCGFTDDEKQGI VDEIVKLSFDNLKNQVKNKGSSVNSKIDNKHFRKGEVRDWNVLYTSEMIXKET 318
```



Db 230 RLAETFGVPTTBEKEGVIEIILKCSFENLSLEVNSKSG---NSKGFLPIENRLYPRK 286  
QY 319 GEVRDWANYLTSEMIMKLE 337  
Db 287 AKDGDKNVFTDMEKID 305

## RESULT 10

US-10-259-165-114  
; Sequence 114, Application US/10259165  
; Publication No. US2003013588A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Tong  
; APPLICANT: Wang, Xun  
; APPLICANT: Chang, Hur-song  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiaki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Moughamer, Todd  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Rickel, Darrell  
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING  
; FILE REFERENCE: 70030-NP  
; CURRENT APPLICATION NUMBER: US/10/259,165  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: US 60/368,327  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26  
; NUMBER OF SEQ ID NOS: 782  
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta  
; SEQ ID NO 114  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-259-165-114

Query Match 24.4%; Score 501.5; DB 14; Length 328;  
Best Local Similarity 38.3%; Pred. No. 1.7e-38;  
Matches 119; Conservative 39; Mismatches 106; Indels 47; Gaps 8;  
QY 48 VSSLPNDYMGD-TMRLYKGFWMQGYLVPGIMAFEDNFKARETDIILTLPKAGTTWTKA 106  
Db 25 ISSLPLETRCAPPFLRQHGFWLPETFLPGLEAARARFPRPSDVLLASFPSKSGTTWLKA 84  
QY 107 LTFALTRDYNHPSSTHPLLFNPHSCYQNLLEYLYMGRENTMPDLDMLNESPRLFAGHI 166  
Db 85 LAFATLNRAAYPPSGEGHPLRRRGPHDCVQFLSALVSDDMFASL----PSPRLSTHL 140  
QY 167 PYSLLPASVL--KSGTKIINISNRKSTFVSFWKF-----GN---LINPKD 207  
Db 141 PYSLLPEGVKADSSGCRIDYICRDPKDLVLSWMLFTKALGTODGPTNGNKMPLSN--- 197  
QY 208 LLDLEKSVDFASGISFCGFENFQAEF-TNAASTNSNLLLSYEEMLEKPVENVKLAELAE 266  
Db 198 -----GPYRHRVLEYWAESKRRPKQVLFRLYEEMTRETTSNVKRLAE 239  
QY 267 FMCGGFTDDBEKQIGVDEIVKLCSPFNKQVKNKNGSSYNSKIDNKHFFRKGVRDWN 326  
Db 240 FMGCPFSGEEADGVDPDAIVGLCSFDHLRSLEVNRNGAN-DFNIKNDSPYRKGVAAGDWN 298  
QY 327 YLTSEMIMKLE 337  
Db 299 YLSPENAAQLD 309

## RESULT 11

US-10-259-165-444  
; Sequence 444, Application US/10259165  
; Publication No. US2003013588A1  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Tong  
; APPLICANT: Wang, Xun  
; APPLICANT: Chang, Hur-song  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiaki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Moughamer, Todd  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Rickel, Darrell  
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING  
; FILE REFERENCE: 70030-NP  
; CURRENT APPLICATION NUMBER: US/10/259,165  
; CURRENT FILING DATE: 2002-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: US 60/368,327  
; PRIOR FILING DATE: 2002-03-27  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26  
; NUMBER OF SEQ ID NOS: 782  
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta  
; SEQ ID NO 444  
; LENGTH: 328  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-259-165-444

Query Match 24.4%; Score 501.5; DB 14; Length 328;  
Best Local Similarity 38.3%; Pred. No. 1.7e-38;  
Matches 119; Conservative 39; Mismatches 106; Indels 47; Gaps 8;  
QY 48 VSSLPNDYMGD-TMRLYKGFWMQGYLVPGIMAFEDNFKARETDIILTLPKAGTTWTKA 106  
Db 25 ISSLPLETRCAPPFLRQHGFWLPETFLPGLEAARARFPRPSDVLLASFPSKSGTTWLKA 84  
QY 107 LTFALTRDYNHPSSTHPLLFNPHSCYQNLLEYLYMGRENTMPDLDMLNESPRLFAGHI 166  
Db 85 LAFATLNRAAYPPSGEGHPLRRRGPHDCVQFLSALVSDDMFASL----PSPRLSTHL 140  
QY 167 PYSLLPASVL--KSGTKIINISNRKSTFVSFWKF-----GN---LINPKD 207  
Db 141 PYSLLPEGVKADSSGCRIDYICRDPKDLVLSWMLFTKALGTODGPTNGNKMPLSN--- 197  
QY 208 LLDLEKSVDFASGISFCGFENFQAEF-TNAASTNSNLLLSYEEMLEKPVENVKLAELAE 266  
Db 198 -----GPYRHRVLEYWAESKRRPKQVLFRLYEEMTRETTSNVKRLAE 239  
QY 267 FMCGGFTDDBEKQIGVDEIVKLCSPFNKQVKNKNGSSYNSKIDNKHFFRKGVRDWN 326  
Db 240 FMGCPFSGEEADGVDPDAIVGLCSFDHLRSLEVNRNGAN-DFNIKNDSPYRKGVAAGDWN 298  
QY 327 YLTSEMIMKLE 337  
Db 299 YLSPENAAQLD 309

## RESULT 12

US-09-854-122-18  
; Sequence 18, Application US/09854122  
; Patent No. US20020016980A1  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, ROBERT  
; APPLICANT: ALBERTE, RANDALL S.  
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
; FILE REFERENCE: PHA-007.01  
; CURRENT APPLICATION NUMBER: US/09/854,122

```
; CURRENT FILING DATE: 2001-09-10
; PRIOR APPLICATION NUMBER: 60/202,529
; PRIOR FILING DATE: 2000-05-10
; NUMBER OF SEQ ID NOS: 51
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 18
; LENGTH: 324
; TYPE: PRT
; ORGANISM: Brassica napus
US-09-854-122-18

Query Match
Best Local Similarity 24.3%; Score 500.5; DB 9; Length 324;
Matches 117; Conservative 58; Mismatches 114; Indels 19; Gaps 10;

QY 37 DSKMYKRYEIVSSPSNDY-GDTRLYKGFQWQVGLVPGIMAFEDNFKARETDIILTT 95
DB 13 DENLQKTKYKDISSLPSEKGLVCMQYQFQRWTKALQGLTQCXHFPAKSDIILVT 72
QY 96 LPKAGTTWTKALTFAILTRDNNHPSPPHLLFFNPHSCVQN-LEYLYMGRE-NTWPDLD 153
DB 73 NPKSGTTWTKALVFAL----INRHKFPVYSVIL---SCYQSALLVPFLGRSLRSPDFD 125
QY 154 MLN-ESPLFAGHIPYSLLPASVLKSGTKIINISRNKSTFVSGFWKGNLINDPKLLD-- 210
DB 126 FSQLSPLMNTSHLSLPSKSSCKIYCCNPKDMFVSLWHEGKLAPETADYP 185
QY 211 LKSVVDIFASGISFCGPPWNNFQAEFTNAASTNSN-LILLSYEEMLEKPVENVKLAFFWG 269
DB 186 IEKAVEAFCOGKFIGGPPWDHVLVYVYASLENPNKVLFSVEEPPKKTGTETIKRIAEFLG 245
QY 270 COFTDDEKQKQIVDIBIVKLCFDMLNKQOVNKGSSVNSKIDNKHPRKGEVDRWANYLT 329
DB 246 COLVGEER----VRAIVKLCFSLSLEVNREG-KLPDSGNETFAFRKGEVGGWRTLT 300
QY 330 SEMIKKLE 337
DB 301 ESIAEVID 308

RESULT 13
US-10-195-144-79
; Sequence 79, Application US/10195144
; Publication No. US20030126646A1
; GENERAL INFORMATION:
; APPLICANT: BROWN, GREGORY G.
; APPLICANT: FORMANOVA, NATASA
; APPLICANT: DENDY, CHARLES
; APPLICANT: LANDRY, BENOIT S.
; APPLICANT: CHEUNG, WING
; APPLICANT: JIN, HUA
; TITLE OF INVENTION: NUCLEAR FERTILITY RESTORER GENES AND METHODS OF USE IN
; FILE REFERENCE: 16313-0136
; CURRENT APPLICATION NUMBER: US/10/195,144
; PRIOR FILING DATE: 2002-10-01
; PRIOR APPLICATION NUMBER: 60/305,026
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/305,363
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 60/308,736
; PRIOR FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 128
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Raphanus sativum
US-10-195-144-79

Query Match
Best Local Similarity 23.5%; Score 483; DB 14; Length 688;
Matches 107; Conservative 55; Mismatches 107; Indels 14; Gaps 7;

QY 56 YWGTMLYKGFQWQVGLVPGIMAFEDNFKARETDIILTTLPKAGTTWTKALTFAILTRD 115
DB 78 FFGGKLYFSILAGNYIFPVLRGFG----PQDQVIAIYSPKSGTTWTKALTVALERS 133
QY 116 VNHPSSTPHLLFPNPHSCVQNLEYLYMGRENTMPDMLNBSPPRLFAGHIPYSLLPASV 175
DB 134 KQHSSD--HPLLYNPHGIIPFLE-IDVYHSSSNLAKFSAPPLRFSHTMPLHIHEAL 190
QY 176 LKSGTKIINISRNKSTFVSFWKFG---NLINPDKLLDLEKSVDFIPASGISFCGPEWNFQ 232
DB 191 KHSCKIIVYCRNVKOTLISCFYSCAIYKIEPTRRV-LESFNEFCGDTNFGFPWDHL 249
QY 233 AEFTNAASTN-SNILLISYEEMLEKPVENVKLAFFWGCGFTDDEKQKQIVDIBIVKLCF 291
DB 250 LSYWRGSLDEPKHVLFWRYEEMKAEPRDQIKRLADFLGCGFTKQBEDSSVSGILDLSL 309
QY 292 DNLKQOVNKGSSVNSKIDNKHPRKGEVDRWANYLTSEMIX 334
DB 310 RNLSSLEANKTGTINN--VEHKFFFRKGEVGGSKNYLTSEMEX 350

RESULT 14
US-10-345-072-79
; Sequence 79, Application US/10345072
; Publication No. US20030237112A1
; GENERAL INFORMATION:
; APPLICANT: BROWN, GREGORY G.
; APPLICANT: FORMANOVA, NATASA
; APPLICANT: DENDY, CHARLES
; APPLICANT: LANDRY, BENOIT S.
; APPLICANT: CHEUNG, WING
; APPLICANT: JIN, HUA
; APPLICANT: LAI, FANG MING
; APPLICANT: LEFOREST, MARTIN
; TITLE OF INVENTION: NUCLEAR FERTILITY RESTORER GENES AND METHODS OF USE IN
; FILE REFERENCE: 16313-0210
; CURRENT APPLICATION NUMBER: US/10/345,072
; PRIOR FILING DATE: 2003-01-16
; PRIOR APPLICATION NUMBER: PCT/US02/22217
; PRIOR FILING DATE: 2002-07-12
; PRIOR APPLICATION NUMBER: 60/305,026
; PRIOR FILING DATE: 2001-07-12
; PRIOR APPLICATION NUMBER: 60/305,363
; PRIOR FILING DATE: 2001-07-13
; PRIOR APPLICATION NUMBER: 60/308,736
; PRIOR FILING DATE: 2001-07-30
; NUMBER OF SEQ ID NOS: 179
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 79
; LENGTH: 688
; TYPE: PRT
; ORGANISM: Raphanus sativum
US-10-345-072-79

Query Match
Best Local Similarity 23.5%; Score 483; DB 15; Length 688;
Matches 107; Conservative 55; Mismatches 107; Indels 14; Gaps 7;

QY 56 YWGTMLYKGFQWQVGLVPGIMAFEDNFKARETDIILTTLPKAGTTWTKALTFAILTRD 115
DB 78 FFGGKLYFSILAGNYIFPVLRGFG----PQDQVIAIYSPKSGTTWTKALTVALERS 133
QY 116 VNHPSSTPHLLFPNPHSCVQNLEYLYMGRENTMPDMLNBSPPRLFAGHIPYSLLPASV 175
DB 134 KQHSSD--HPLLYNPHGIIPFLE-IDVYHSSSNLAKFSAPPLRFSHTMPLHIHEAL 190
QY 176 LKSGTKIINISRNKSTFVSFWKFG---NLINPDKLLDLEKSVDFIPASGISFCGPEWNFQ 232
DB 191 KHSCKIIVYCRNVKOTLISCFYSCAIYKIEPTRRV-LESFNEFCGDTNFGFPWDHL 249
QY 233 AEFTNAASTN-SNILLISYEEMLEKPVENVKLAFFWGCGFTDDEKQKQIVDIBIVKLCF 291
```

Job time : 35 secs

Db 250 LSYWGSLEDKXVLFMEYENKAEPRQIKRLADFLGCPFTKQEESSGSDGILDCSL 309  
QY 292 DNLKQVQVNGSSYNSKIDNKHFFPKGEVDMANYLTSEMIX 334  
Db 310 RNLSSLEANKTGTINN--VEHKFFPKGEVDSKNYLTSEMEX 350

RESULT 15  
US-10-259-194A-330  
; Sequence 330, Application US/10259194A  
; Publication No. US20040010815A1  
; GENERAL INFORMATION:  
; APPLICANT: Lange, Markus B.  
; APPLICANT: Ghassseman, Majid  
; APPLICANT: Briggs, Steven P.  
; APPLICANT: Cooper, Bret  
; APPLICANT: Glazebrook, Jane  
; APPLICANT: Goff, Stephen A.  
; APPLICANT: Katagiri, Fumiaki  
; APPLICANT: Kreps, Joel  
; APPLICANT: Moughamer, Todd  
; APPLICANT: Provart, Nicholas  
; APPLICANT: Ricke, Darrell  
; APPLICANT: Zhu, Tong  
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES  
; FILE REFERENCE: 70029-NP  
; CURRENT APPLICATION NUMBER: US/10/259,194A  
; CURRENT FILING DATE: 2003-01-07  
; PRIOR APPLICATION NUMBER: US 60/325,277  
; PRIOR FILING DATE: 2001-09-26  
; PRIOR APPLICATION NUMBER: US 60/370,743  
; PRIOR FILING DATE: 2002-04-04  
; PRIOR APPLICATION NUMBER: US 60/370,620  
; PRIOR FILING DATE: 2002-04-04  
; NUMBER OF SEQ ID NOS: 662  
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta  
; SEQ ID NO 330  
; LENGTH: 343  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
US-10-259-194A-330

Query Match 23.3%; Score 479; DB 15; Length 343;  
Best Local Similarity 37.9%; Pred. No. 2,4e-36;  
Matches 125; Conservative 54; Mismatches 121; Indels 30; Gaps 14;  
QY 18 GILALEKCFGSKNEQKEEDSKMYKRYREIVSLPSNDYWDGTMRLYKGFQMGYLVFGI 77  
Db 18 GTLAAAE--ARSPLPKEE-----FGDLVAALPRKEQYLDG-RLYEGFWLPEHYAPGI 66  
QY 78 MAFEDNFKARETDIILTLTPKAGTTWKALTFAILTRDVNHPSPHPLFPNPHSCYQN 137  
Db 67 IAFRRRTPRADVVLIASYPKCGTTLKALAFANMTR-AAYPAD-EHPLRLNPHDVIFF 124  
QY 138 LEVLYM-GRENTMPDLDMLNESPRIFAGHIPYSLLPASVYK----SGTKIINISNRKST 192  
Db 125 VEDVFTDGEH--AKLDML-PSPLRLNTHTPYQLLPESVWAGDGGGCKVYVICRDPKM 180  
QY 193 FVSFWKFGNLIIPDKLLDLEKSDVDFASGISFCGPEWNPQAEFTNAA-STNSNLLILSYE 251  
Db 181 VVSLYHFMERLQPD--LSLAGVVSVDGTVPFPGPMWDHILGYWRASVSRPDRVLLRYE 238  
QY 252 EMLEKEV--ENVKLAEFMGCGFTDDEEKQIGVDEIVKLCSPDNLKQOVNKNKS--SYN 307  
Db 239 DLLRDGAAGEHVRAVAFVGRFSAEEAAGAVASVVELCSFERMKALEVNERGTAGSYK 298  
QY 308 SKIDNKHFFPKGEVDMANYLTSEMIX 337  
Db 299 S-MPRDAFFRKGVAGDWNHMSPTAARLD 327

Search completed: March 20, 2004, 01:40:31

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 20, 2004, 03:02:04 ; Search time 5120 Seconds  
(without alignments)  
3309.996 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059

Sequence: 1 TRGITGIAVACLIIMAGIL.....CLLSYSFVNNFILLKKKK 391

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3470272 seqs, 21671516995 residues

Total number of hits satisfying chosen parameters: 6940544

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+ p2n.model -DRV=xlh  
-Q=/cgn2.1/USPTO.spool/US09854122/runat.18032004.100000.3253/app.query.fasta.1.583  
-De=GenEmbl -CFMT=fastap -SUFFIX=rg -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -STAT=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-DOALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854122 @CGN.1.1.3508 @runat.18032004.100000.3253 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

GenEmbl.\*  
1: gb.ba.\*  
2: gb.htg.\*  
3: gb.in.\*  
4: gb.om.\*  
5: gb.ov.\*  
6: gb.pat.\*  
7: gb.ph.\*  
8: gb.pl.\*  
9: gb.pr.\*  
10: gb.ro.\*  
11: gb.sts.\*  
12: gb.sy.\*  
13: gb.un.\*  
14: gb.vi.\*  
15: em.ba.\*  
16: em.fun.\*  
17: em.hum.\*  
18: em.in.\*  
19: em.mu.\*  
20: em.om.\*  
21: em.or.\*  
22: em.ov.\*  
23: em.pat.\*  
24: em.ph.\*  
25: em.pl.\*  
26: em.ro.\*  
27: em.sts.\*  
28: em.un.\*

29: em.vi.\*  
30: em.htg.hum.\*  
31: em.htg.inv.\*  
32: em.htg.other.\*  
33: em.htg.mus.\*  
34: em.htg.pin.\*  
35: em.htg.rod.\*  
36: em.htg.mam.\*  
37: em.htg.vrt.\*  
38: em.sy.\*  
39: em.htgo.hum.\*  
40: em.htgo.mus.\*  
41: em.htgo.other.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1996	96.9	1192	6	AX338910 Sequence
2	647.5	31.4	111989	2	AX135566 Medicago
C 3	620.5	30.1	88010	8	AT114D3 Arabidops
C 4	620.5	30.1	89214	8	ATF14D17 Arabidops
C 5	594	28.8	138586	8	AC144726 Medicago
C 6	593	28.8	166490	2	AC146862 Medicago
C 7	587	28.5	107440	2	AC119410 Medicago
C 8	579	28.1	138586	8	AC144726 Medicago
C 9	578.5	28.1	102183	8	AC027134 Arabidops
C 10	578.5	28.1	103157	8	AC011810 Arabidops
C 11	574	27.9	108387	8	AC016662 Arabidops
C 12	573	27.8	1347	8	AY087493 Arabidops
C 13	572.5	27.8	1212	8	AY085411 Arabidops
C 14	570	27.7	1221	8	AY042887 Arabidops
C 15	567.5	27.6	113531	2	AC146342 Medicago
C 16	567	27.5	1041	8	AX697959 Sequence
C 17	567	27.5	1072	8	BT005398 Arabidops
C 18	567	27.5	1151	8	AK117408 Arabidops
C 19	567	27.5	1247	8	AK117463 Arabidops
C 20	567	27.5	44879	8	AC036238 Arabidops
C 21	566	27.5	1281	8	AY088081 Arabidops
C 22	566	27.5	108387	8	AC016662 Arabidops
C 23	564.5	27.4	1315	8	AK058698 Oryza sat
C 24	563.5	27.4	1092	8	AY081540 Arabidops
C 25	562.5	27.3	1017	6	AX697955 Sequence
C 26	559	27.1	1084	8	BT004984 Arabidops
C 27	558.5	27.1	1053	6	AX697957 Sequence
C 28	555.5	27.0	1077	6	AX080458 Sequence
C 29	555.5	27.0	1080	8	BT008847 Arabidops
C 30	555.5	27.0	1273	8	AY084999 Arabidops
C 31	555.5	27.0	1347	8	AY099809 Arabidops
C 32	555.5	27.0	86380	8	AB010697 Arabidops
C 33	553.5	26.9	121720	8	AC021044 Arabidops
C 34	549.5	26.7	1270	8	AY054219 Arabidops
C 35	548.5	26.6	1090	8	BT002108 Arabidops
C 36	548.5	26.6	1154	8	AY087921 Arabidops
C 37	548.5	26.6	1216	8	AY092961 Arabidops
C 38	548.5	26.6	87543	8	AC006836 Arabidops
C 39	546	26.5	1994	8	AF000307 Brassica
C 40	545	26.5	996	6	AX506332 Sequence
C 41	545	26.5	996	8	AF325065 Arabidops
C 42	542.5	26.3	1147	8	AY442306 Brassica
C 43	542.5	26.3	90341	8	AC006232 Arabidops
C 44	542.5	26.3	118196	8	AC005824 Arabidops
C 45	542	26.3	1179	8	AF375458 Arabidops

# ALIGNMENTS

RESULT 1

AX338910  
LOCUS AX338910 1192 bp DNA linear PAT 09-JAN-2002  
DEFINITION Sequence 15 from Patent WO0185971.  
ACCESSION AX338910  
VERSION AX338910.1 GI:18129074  
KEYWORDS  
SOURCE Zostera marina  
ORGANISM Zostera marina  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Zosteraceae; Zostera.  
REFERENCE 1  
AUTHORS Alberte, R.S. and Smith, R.D.  
TITLE Transgenic plants incorporating traits of Zostera marina  
JOURNAL Patent: WO 0185971-A 15 NOV-2001;  
Phycogen, Inc. (US)  
FEATURES  
    Location/Qualifiers  
        1..1192  
        /organism="Zostera marina"  
        /mol\_type="unassigned DNA"  
        /db\_xref="taxon:29655"  
        joinr(<1..33,37..1041,1045..1059,1063..1077,1081..1086,  
          1090..1119,1123..>1191)  
        /notes="unnamed protein product"  
        /codon\_start=1  
        /protein\_id="CAD20565.1"  
        /db\_xref="GI:18129075"  
        /translation="TRGITGAVACLELIMAGILALEKCFKSKNEQEKEDSKMYKRY  
          REIVSSLPNSDWGDTWELGYKFWQMGVLPVGINAFEDNFKARETDIILTLPKAGTT  
          WTKALTPAILRDVNHPSPTPLLPFNPHVGVNLEVMGRENTPMDLMLNESPR  
          LPAGHPIYSLPASVLSKGTIKINISRNKSTFVSFWKFGNLINPDKLDELKSVDFI  
          ASGISFCPEWFOHEFNAASTNSNLLLSYEEMLEKPVENVKKLAEFWGCGFTDDE  
          EKQGVIVDVLKCPDNKNOOVNKNSSYNSKIDNKHFFPKGVRVDRWNYLISEMK  
          KLETAGKINSEKELLSNKLNTNLFNLNNYCEKSNISLCLLSYSEVFINNFIILLKKKK"

Alignment Scores:  
Pred. No.: 4,448-161 Length: 1192  
Score: 1996.00 Matches: 391  
Percent Similarity: 98.49% Conservatives: 0  
Best Local Similarity: 98.49% Mismatches: 6  
Query Match: 96.94% Indels: 6  
DB: Gaps: 6

US-09-854-122-16 (1-391) x AX338910 (1-1192)

QY 1 ThrArgGlyIleThrGlyIleAlaValAlaCys---LeuProLeuIleMetAlaGlyIle 19  
DB 1 ACGCGGGGAATAACTGGAATCGTGTGCTTCTAGCTACCCTGATATGCTGGAATT 60  
QY 20 LeuAlaLeuGluLysCysPheGlySerLysAsnGluGlnGluLysGluLysSerLys 39  
DB 61 TTAGCTTTGGAGAAATGTTTCGATCCAGAAATGAGCAGAGAGAGAGATTCCTCAA 120  
QY 40 MetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTrpGlyAsp 59  
DB 121 ATGTACAGAGATATAGAGAGATTTGTTCTTCTCCTCCCTCGAATGATTATTGGGGGAT 180  
QY 60 ThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAla 79  
DB 181 ACCATGAGGTGTACAGGAGATTTTGGCAATGGGATATCTTGTACCTGGTATCATGGCT 240  
QY 80 PheGluAspAsnPhelYsAlaArgGluThrAspIleLeuThrThrLeuProLysAla 99  
DB 241 TTCAAGAGATAATTCAAGGCTCGAGAGAGAGATATCTTACGACTTCTCCAAAGGCT 300  
QY 100 GlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisPro 119  
DB 301 GGAACGACATGAGAGAGGACGCTGAGCTTTGGCATCTTCAACAGAGATTTTAAACCCCA 360  
QY 120 SerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGlu 339  
DB 361 TCATCACCGACACATCCACTTTTGTCTTCAACCCCTCATTCGCTGTGTTCAAAATTTGGAG 420

QY 140 TyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerPro 159  
DB 421 TATTTGTACATGGGTAGAGAAAATACGATGCGACCTCGATATGTTGAATGAATCCCGC 480  
QY 160 ArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGly 179  
DB 481 AGGTGTTTGGCGGACATCCCATACCTTTGTTGCGGCTCTGTTTGAATTCGGA 540  
QY 180 ThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLysPhe 199  
DB 541 ACAAAATCATCAATATTAAGCGCAACCGTAAGAGTACATTTGTCTCTTTTGGAAATTT 600  
QY 200 GlyAsnLeuIleAsnProAspLysLeuAspLeuGluLysSerValAspIlePheAla 219  
DB 601 GCGAATCTGATTACCCCGACAAAGTTATTGGACCTTCGAAAGAGCGTTGATATCTTCGA 660  
QY 220 SerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPheThrAsnAlaAla 239  
DB 661 TCGGGAATCTCTTTTGGACCGCAATGGAATTTCCAAAGCGGAGTTCCCAATTCGCGC 720  
QY 240 SerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGlu 259  
DB 721 TCTACTAATTCAACTTCTGTTGAGTTACGAGAAATGTTTAGAGAGCCATTTGAA 780  
QY 260 AsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluLysGln 279  
DB 781 AATGTGAAGAACTAGCTGAGTTTCATGGGATGCGGTTCCACAGCGATCAGGAGAAACAA 840  
QY 280 GlyIleValAspGluIleValLysCysSerPheAspAsnLeuLysAsnGlnVal 299  
DB 841 GGGATTTGTTGATGAGATAGTTAAACTTTGTAGCTTCGCAATCTGAAGAAATCAACAGGTG 900  
QY 300 AsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHisPhePheArgLysGly 319  
DB 901 AACAAAACGGATCAAGCTACATTCGAAATCGACACCAAGCAATTTCTTCAGGAAGGT 960  
QY 320 GluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGluThrAla 339  
DB 961 GAGGTGAGAGATTTGGCAAACTATCTAACGTCGGAATGATTAAGAAATCTGAGACGGCC 1020  
QY 340 GlyLysIleAsnGluSerGlu---LysHisLeuLeuSer---AsnLysAsnLeuThr--- 356  
DB 1021 GGAATAATTAATGATCAGATTAAGCAATTTATCTCTGAAATTAAGAAATCTTACATGA 1080  
QY 357 AsnPhe---AsnLeuAsnAsnTyrCysGluLysSerAsn---IleSerLeuCysLeuLeu 374  
DB 1081 AACTTCTGAAATCTTAATAATTACTGTGAGAAATCGAACTAAATATCTCTTGTGTTATTA 1140  
QY 375 SerTyrSerPheValIleAsnAsnPhelLeuLeuLysLysLysLysLys 391  
DB 1141 TCGTATTTCATTCGTAATAATAATTTTCATTTTGTAAAAAATAAAAAA 1191

RESULT 2  
AC135566 111989 bp DNA linear HTG 12-DEC-2003  
LOCUS Medicago truncatula clone mth2-22g8, WORKING DRAFT SEQUENCE, 3  
DEFINITION ordered pieces.  
ACCESSION AC135566  
VERSION AC135566.13 GI:39752715  
KEYWORDS HTG; HTGS PHASE2; HTGS DRAFT.  
SOURCE Medicago truncatula (barrel medic)  
ORGANISM Medicago truncatula  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
Medicago.  
REFERENCE 1 (bases 1 to 111989)  
AUTHORS Shaul, S., Lin, S., Dixon, R., May, G., Sumner, L., Gonzales, B.,  
Cook, D., Kim, D. and Roe, B.A.  
TITLE Medicago truncatula BAC Clone mth2-22g8  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 111989)

AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (18-OCT-2002) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

REFERENCE 3 (bases 1 to 111989)

AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D. and Roe,B.A.

TITLE Direct Submission

JOURNAL Submitted (12-DEC-2003) Department Of Chemistry And Biochemistry, The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman, OK 73019, USA

COMMENT ON Dec 12, 2003 this sequence version replaced gi:39573775.  
----- Genome Center  
Center: Department Of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR  
-----  
\* NOTE: This is a 'working draft' sequence. It currently consists of 3 contigs. Gaps between the contigs are represented as runs of N. The order of the pieces is believed to be correct as given, however the sizes of the gaps between them are based on estimates that have been provided by the submitter.  
\* This sequence will be replaced  
\* by the finished sequence as soon as it is available and the accession number will be preserved.  
\* 1 11239: contig of 11239 bp in length  
\* 11240 11339: gap of unknown length  
\* 11340 20610: contig of 9271 bp in length  
\* 20611 20710: gap of unknown length  
\* 20711 111989: contig of 91279 bp in length.  
FEATURES  
source location/Qualifiers  
1..111989  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3880"  
/clone="mtn2-22g8"  
/clone\_lib="Medicago truncatula BAC library H2"

ORIGIN

Alignment Scores:

Pred. No.:	1,01e-43	Length:	111989
Score:	647.50	Matches:	126
Percent Similarity:	60.00%	Conservative:	74
Best Local Similarity:	38.86%	Mismatches:	125
Query Match:	31.45%	Indels:	15
DB:	2	Gaps:	9

US-09-854-122-16 (1-391) X AC135566 (1-111989)

Qy 31 GluGlnGlyLysGluAlaSpSerLysMetTyLysArgTyArgGluIleValSerSer 50  
Db 43180 GATGAATCCATTGAAGACAAGGTCAACTTAOCCAAGAAAACAAGAATACTTCTTTCT 43239

Qy 51 LeuProSerAsnAspTyTrpGlyAepThr---MetArgLeuTyLysGlyPheTrpGln 69  
Db 43240 CTTCCAAGACAGAAAGTTGGAGAACACCTTATATTATTCATTCCAAGATTTTGGTGC 43299

Qy 70 MetGlyTyTrpLeuValProGlyIleMetAlaPheGluAepAsnPhelyAlaargGluThr 89  
Db 43300 CAACCAGCTGAATCCAAGCCATAACCATTTCCAAAAGCATTTCCAAGCTAAGAAAGT 43359

Qy 90 AspIleIleLeuThrThrLeuProLysAlaGlyThrTrpThrLysAlaLeuThrPhe 109  
Db 43360 GATGTTTTTGTC AACCGTACCAAATCAGGCACAACTTGGTTAAAGCTCTTACCTAT 43419

Qy 110 AlaIleLeuThrArgspValasnHisProSerSerProThrHisProLeuLeuPhePhe 129  
Db 43420 GCCATTATGAATGCCCAACCAACATTTTCATTATCCAAAAACATCCCTTTGTTAGTTTC 43479

Qy 130 AsnProHisSerCysValGlnAsnLeuGluTyLeuty-MetGlyArgGluAsnThrMet 149

Coordinator: Marcel Salanoubat and Francis Quetier, Grouperment d'Interet Public, Centre National de Sequenage - GENOSCOPE; 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France;

<http://www.genoscope.cns.fr>

Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thai/>.

#### FEATURES

source Location/Qualifiers

1. 88010

/organism="Arabidopsis thaliana"

/mol\_type="genomic DNA"

/variety="Columbia"

/db\_xref="taxon:3702"

/chromosome="3"

1145. 2116

/gene="Tl14D3.10"

/complement(1145. 2116)

/gene="Tl14D3.10"

/complement(1145. 2116)

/gene="Tl14D3.10"

/note="strong similarity to FLAVONOL 4'-SULFOTRANSFERASE - Flaveria chloraefolia, EMBL:M84136"

/codon\_start=1

/product="sulfotransferase-like protein"

/protein\_id="CAB72145.1"

/db\_xref="GI:6911845"

/db\_xref="GOA:Q9M1V2"

/db\_xref="SPTREMBL:Q9M1V2"

/translation="MENNRIEDLNERTKTLISLPSDKDFTGKTKICKYQGQWYTHNV LQAVINQKFKPDQTDIIIVASPEKCGTTLKALTALHRSKQPSHDDHPLSNRP HVLVPYFEIDLRLSENPDLTKSSSPLESTHVPSTLQELGKSTCKIVTSRVK DTLVYHFFKQTDKTLISSEFDTFEMFCRGVSFGPTWDHVLVYWGSLDPNHV LFMKFEKPEPDIQFAFLGCPFTKEEESGVSDEIIDLCSLNLSSLEIKTGL KLSGRNKMFRKGEVGMKNTLTPMENKIDMIQELQNSGLKF"

/complement(1145. 2116)

/gene="Tl14D3.10"

/number=1

3483. 4472

/gene="Tl14D3.20"

/complement(3483. 4472)

/gene="Tl14D3.20"

/complement(3483. 4472)

/gene="Tl14D3.20"

/note="strong similarity to FLAVONOL 4'-SULFOTRANSFERASE - Flaveria chloraefolia, EMBL:M84136"

/codon\_start=1

/product="sulfotransferase-like protein"

/protein\_id="CAB72146.1"

/db\_xref="GI:6911846"

/db\_xref="GOA:Q9M1V1"

/db\_xref="SPTREMBL:Q9M1V1"

/translation="MDEKKITWVNDLSSESKTLISLPSDKNSTGVNCKYQCGW VTPPIQGVLPQRKFKPDQTDIIIVASPEKCGTTLKALTALVRRSKHPSHDDHPL LSDNPHVLSPLMXYLCSNPDLTKFSSSRLESTHMPSTLQELGKSTCKIVM SRNKDTLVSYHFFCKQTDNIISSVEDTFEMFCRGVNFGEFWDHVLVYWGSL EDPNVLFMKFEKPEPDIQFAFLGCLFTKEEESGLVDEIIDLCSLNLSSLE INKTKLHSTGRNKTFKRGVGMKNTLTPMENKIDMIQELQNSGLKF"

/complement(3483. 4472)

/gene="Tl14D3.20"

/number=1

5519. 9216

/gene="Tl14D3.30"

/complement(5519. 9216)

7034. 7426,7615. 7696,8280. 8578,9067. 9216))

/gene="Tl14D3.30"

/complement(5519. 9216)

7034. 7426,7615. 7696,8280. 8578,9067. 9216))

/gene="Tl14D3.30"

/note="strong similarity to 2-phosphoglycerate kinase - Methanococcus Jannaschii, PIR:A64485

Contains ATP/GTP-binding site motif A (P-loop)

A164-171:Prokaryotic membrane lipoprotein lipid

attachment site A158-168:Prokaryotic membrane lipoprotein

lipid attachment site AA386-396"

/codon\_start=1

/product="putative protein"

/protein\_id="CAB72147.1"

/db\_xref="GI:6911847"

/db\_xref="SPTREMBL:Q9M1V0"

/translation="MWEATKVLIVVREEGDDDDNDDDSFRVTRPVLOSTLOLMGCK ARHAFKLSRRVVELIRSEGCSTSPENGKPEPAKEVGGSTCVKLNCLVAGDVKN KSPFENYKRTTIVSREIFDVVCDALAEYIKVGRDQADLILACIRERKESVTV LDCPTGCGKSTISALIGSLGITVTVSTDSIRHMRSPFADEKONPULWASTTHAGY LDPVAEASKAKKAKKLSGRVGNAGNAQTDAGSNSTTELLSHKQMALEGYKQSE MVIDSLDLTLTWEERNESVVVEGHLNLFVMLMKKHPISIVPFMYVIANEEKHLER FAYPAKYMTLDPAKKNKYKIRNRTIQDYLCRADKHLVPKNNTNVDKSVATIHAT VFGCLRRREGEKLYDITNTVSVIDDEHNOCAANSLTSKGMFOVIOGSSRRFWA LCNTDGTAKTWPVAVSGKIKVPVNTMEDDGTGHOHKAEPVNLQFHFGISAWPSD GATSHAGSVDDLADIILETSGRHYSSCCSPRTSDSGSKELMEEQSVNGSDEDEED DDFHEDPSDELDNDRNDRDEIGSVDEESTSDDEBYDDLMEQSVNGSDEDEEDER DTISMVONNHNEASTKNDKYSQNLDFLKTNQPLTESLELTSYRNRMGVAASD KAKWRKRSLSIPPVKGKHSIIDQILANQTDVSL"

/complement(5519. 5902)

/gene="Tl14D3.30"

/number=1

/complement(5903. 6006)

/gene="Tl14D3.30"

/number=1

/complement(6007. 6129)

/gene="Tl14D3.30"

/number=2

/complement(6130. 6216)

/gene="Tl14D3.30"

/number=2

/complement(6217. 6936)

/gene="Tl14D3.30"

/number=3

/complement(6937. 7033)

/gene="Tl14D3.30"

/number=3

/complement(7034. 7426)

/gene="Tl14D3.30"

/number=4

/complement(7427. 7614)

/gene="Tl14D3.30"

/number=4

/complement(7615. 7696)

/gene="Tl14D3.30"

/number=5

/complement(7697. 8279)

/gene="Tl14D3.30"

/number=5

/complement(8280. 8578)

/gene="Tl14D3.30"

/number=6

/complement(8579. 9066)

/gene="Tl14D3.30"

/number=6

/complement(9067. 9216)

/gene="Tl14D3.30"

/number=7

14402. 14473

/note="72 bp long terminal repeat"

14495. 14861

/note="similarity to fragments of copia-like polyprotein"

19002. 19073

/note="72 bp long terminal repeat"

19415. 21625

/gene="Tl14D3.40"

20930. 21066,21169. 21272,21360. 21450,21533. 21625)

/gene="Tl14D3.40"

/note="strong similarity to n-acetylglucosaminyl-phosphatidylinositol - Schizosaccharomyces pombe, PIR:140367"

/codon\_start=1

/product="n-acetylglucosaminyl-phosphatidylinositol-like protein"  
 /protein\_id="CAB72148.1"  
 /db\_xref="GI:5911848"  
 /db\_xref="GOA:Q9MIU9"  
 /db\_xref="SPTREMBL:Q9MIU9"  
 /translation="MAEPPKRLVLMVSDFFEFNFGGVENHIVYLSQCILLKLGHKVVMV  
 THAVGNSGVRYVTGGLKYVYVPRPFWMTTPTVYGTLPVSTLPIVREKLTIVVGH  
 QAFSLCEALMARTINGYKVFVTDHSLGFADVGSINHNKVLQFSLADIDQAICVSH  
 TSKENTVLRGLSPAKVFMIPNAVDITAMFPAKSVRPSDITITIVVIRLVTRKGADILL  
 VEVPTECFYPNVFRVGGDGPKHVRLEEMREKESLQDRVEMLGAVPHGRVRSVLVT  
 CHIFNLSLEAFCAILAEASCGLLTVSTRVGGFLHGLQVLIPDMVVLAEPPDDMV  
 RAIEKASILPTTPNPEMHNRMKLYSMQDVAKTEIVYDRALKCSNRSLLERLMRFL  
 SCGAWAGKLCFQWIMILDYLLWRLQLLPQDEIDIEAPDLCLECHRGVEVSEGLRKKIK  
 "
 exon 19415..19534 /gene="T14D3.40" /number=1  
 intron 19535..19837 /gene="T14D3.40" /number=1  
 exon 19838..20410 /gene="T14D3.40" /number=2  
 intron 20411..20507 /gene="T14D3.40" /number=2  
 exon 20508..20662 /gene="T14D3.40" /number=3  
 intron 20663..20757 /gene="T14D3.40" /number=3  
 exon 20758..20837

Alignment Scores:  
 Pred. No.: 1,54e-41 Length: 88010  
 Score: 620.50 Matches: 143  
 Percent Similarity: 37.55% Conservative: 78  
 Best Local Similarity: 37.24% Mismatches: 126  
 Query Match: 30.14% Indels: 37  
 DB: 8 Gaps: 11

US-09-854-122-16 (1-391) x ATT14D3 (1-88010)

QY	25	CysPheGlySerLysAsnGluGlnGluLysGluGlu-AspSerLysMetTyrLysArgTy	44
Db	4476	TGTAATGATCAGAAAAAGATTACGATGAACGTGCGAAACGATGATTAAGCGAAGATC	4417
QY	44	xArgGluIleValSerLeuProSerAsn--AspTyrTrpGlyAspThrMetArgLe	63
Db	4416	AAAAACTCTATCTCTTCACCTCTCTTCAGACAAAANTCCATCGGGTAATGTTGTAA	4357
QY	63	uTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetalpheGluAspAs	83
Db	4356	GTACCAAGGATGTTGGTATCTCCGCCCTATCTTCGAAGGTGTCCTCAATTTCCAGAAA	4297
QY	83	nPheLysAlaArgGluThrAspIleLeuThrThrLeuProLysAlaGlyThrThrTr	103
Db	4296	CTTTAAGCCTCAAGACACCGNATCATCGTTCGTTCTCCCTTAATCGGCACCATG	4237
QY	103	pThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisProSer--SerPr	122
Db	4236	GCTCAAGCGCTTACATTCGCACCTCGTTCGTAGATCG--AAGACCCCTTCTCATGATGA	4190
QY	122	oThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGlu--TyrLe	141
Db	4179	TCATCATCATCTTTCTTGATATCCACAGTCTCTTCACCGCTCCCTTGAGATGATCT	4120
QY	141	uTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLe	161
Db	4119	CTATCTATGTAGCGAAAT-----CCGACCTTACCAAGTTCTCATCATCTTCAGGCT	4066



Biochemie, Am Klopferspitz 18a, D-92152 Martinsried, FRG, E-mail: lenckemips.biochem.mpg.de, mayer@mips.biochem.mpg.de  
Coordinator: Marcel Salanoubat and Francis Quetier, Groupement d'Interet Public, Centre National de Sequencage - GENOSCOPE; 2 rue Gaston Cremieux, BP191, 91006 Evry Cedex, France;  
<http://www.genoscope.cns.fr>  
Information on performance of analysis and a more detailed annotation of this entry and other sequences of chromosomes 3, 4 and 5 can be viewed at: <http://www.mips.biochem.mpg.de/proj/thal/>.  
Location/Qualifiers  
1. .89214  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/variet="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="3"  
5576. .6238  
/gene="F14D17\_10"  
complement(5576. .6238)  
/gene="F14D17\_10"  
complement(5576. .6238)  
/gene="F14D17\_10"  
note="similarity to predicted protein, Arabidopsis thaliana"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB89309.1"  
/db\_xref="GI:7671395"  
/db\_xref="SPTREMBL:Q9LX11"  
/translation="MELQKILVAHEELRKDEQLIHLEDVLTITLKERDEALEKYNH LLNLLIQQQQONQKQELVTPPLSGASSIIEDEGVQQQLNSKSFSSDTE RSIKSPVDPVNNQIQVEGDEMLATLPKPEKLLQAVIKAGPLLQYLLIA GPLQWRHPHPPLTSEIPTVPIPLPQNNCGGNSKKRAFSIDETSETKYQKVL LH"  
complement(5576. .6238)  
/gene="F14D17\_10"  
/number=1  
12760. .13001  
/gene="F14D17\_20"  
join(12760. .12885,12909. .13001)  
/gene="F14D17\_20"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB89310.1"  
/db\_xref="GI:7671396"  
/db\_xref="SPTREMBL:Q9LX10"  
/translation="MTTATVAPTTVVASSGDGCGSGDNGNPNNGVAVPG NSGGGGGDDGDDGGDDGGGGVWLW"  
12760. .12885  
/gene="F14D17\_20"  
/number=1  
12886. .12908  
/gene="F14D17\_20"  
/number=1  
12909. .13001  
/gene="F14D17\_20"  
/number=2  
18756. .18927  
/product="tRNA-Gly"  
/note="tRNA predict as a tRNA- Gly : anticodon tcc"  
19510. .20866  
/gene="F14D17\_30"  
join(19510. .19614,20074. .20142,20264. .20387,20469. .20529,20602. .20866)  
/gene="F14D17\_30"  
/codon\_start=1  
/product="putative protein"  
/protein\_id="CAB89311.1"  
/db\_xref="GI:7671397"  
/db\_xref="SPTREMBL:Q9LXH9"  
/translation="MRQSLSDITNSQBELNLHIGSPSGHVRMLKVKALHEVT CKNALKACDCEKCKRNALTDESPPKPFVPTRRPFRSLG ASTTNQNEAKVSKSEKRRHLRRSPRVSAAQEVTENLFEIEDIQLITPNDMSQDND

LTMSHTRKKEEGEGLRTRHDFEIVRSPVDRSLRRAERIHSYKG"  
19510. .19614  
/gene="F14D17\_30"  
/number=1  
19615. .20073  
/gene="F14D17\_30"  
/number=1  
20074. .20142  
/gene="F14D17\_30"  
/number=2  
20143. .20263  
/gene="F14D17\_30"  
/number=2  
20264. .20387  
/gene="F14D17\_30"  
/number=3  
20388. .20468  
/gene="F14D17\_30"  
/number=3  
20469. .20529  
/gene="F14D17\_30"  
/number=4  
20530. .20601  
/gene="F14D17\_30"  
/number=4  
20602. .20866  
/gene="F14D17\_30"  
/number=5  
24158. .26563  
/gene="F14D17\_40"  
join(24158. .24366,24453. .24777,24884. .25036,25129. .25377,25629. .25721,25837. .25843,25968. .26074,26339. .26563)  
/gene="F14D17\_40"  
note="strong similarity to cytochrome P450 d13695c, Arabidopsis thaliana, PIR:C71417"  
/codon\_start=1  
/product="cytochrome P450-like protein"  
/protein\_id="CAB89312.1"  
/db\_xref="GI:7671398"  
/db\_xref="GOA:Q9LXH8"  
/db\_xref="SPTREMBL:Q9LXH8"  
/translation="MNLFWNTGCVILVVARVGHVWYQWSPKNGKLPQSGMGFPI IGTELDFPYGYFSEISPYLKKKRLRYGLFNTILGVTVVSTDKDVMNEILRQENK SFILSYDGLMKPLGKDSLFIHGHKIQTILHLSSSEGLKRLKMDRVTREH LSSKAKTGLDVKDASKLI IAHLPKMSNLKPTQAKLGI KFAFTDFWTSYLI SAGKGIYNTLWACRGEMREIKDLYMRKTSSEKYDDPLNTAIESEKAGELLNENAI I TLPTSCVTQDTSKAJGLAVKELLEKVAELKKEHEVILESEKDEGGVWBEY RHKMTFTNKSGLTIPAGINIVMLIPSVHFDPEIYENPEFNPMWEGELAGSKTF MVEGTGLRQCAGAEFARLQISVFLHLVTTINFLHQCCEVLRVPAHLPNGISINIS KCPKK"  
24158. .24366  
/gene="F14D17\_40"  
/number=1  
24367. .24452  
/gene="F14D17\_40"  
/number=1  
24453. .24777  
/gene="F14D17\_40"  
/number=2  
24778. .24883  
/gene="F14D17\_40"  
/number=2  
24884. .25036  
/gene="F14D17\_40"  
/number=3  
25037. .25128  
/gene="F14D17\_40"  
/number=3  
25129. .25377  
/gene="F14D17\_40"  
/number=4  
25378. .25628  
/gene="F14D17\_40"

```

exon      /number=4
25629..25721
/genes="F14D17_40"
intron    /number=5
25722..25836
/genes="F14D17_40"
exon      /number=5
25837..25843
/genes="F14D17_40"
intron    /number=6
25844..25967
/genes="F14D17_40"
exon      /number=6
25968..26074
/genes="F14D17_40"
intron    /number=7
26075..26338
/genes="F14D17_40"
exon      /number=7
26339..26563
/genes="F14D17_40"
gene       /number=8
32738..33813
/genes="F14D17_50"
CDS        /genes="F14D17_50"
33886..338813
/codon start=1
/product="putative protein"
/protein_id="CAB89413.1"
/db_xref="GI:7671399"
/db_xref="SPTREMBL:Q9LXH7"
/translation="MAELSSGKSTLASIANWYOFARKIBETDFLLSPNSYKDSI
DYTSLELRARLRWDRDAIAADICNARVKGDSQNGKIEPKWSRKIHPLLSNR
KYWAENLRRLRQDSDIAEKSISGKSDSRVEREVPPTPASFSSTGVGEIRDSEFF

Alignment Scores:
Pred. No.: 1 56e-41      Length: 89214
Score: 620.50      Matches: 143
Percent Similarity: 57.55%      Conservative: 78
Best Local Similarity: 37.24%      Mismatches: 126
Query Match: 30.14%      Indels: 37
DB: 8      Gaps: 11

US-09-854-122-16 (1-391) x ATF14D17 (1-89214)

Qy 25 CysPheGlySerLysAenGluGlnGluLysGluGlu-AspSerLysMetTyrLysArgTy 44
Db 81424 TGTAAATGATGAGAAAAGAGATTAGCAACGTGAGAGAACGATGAGTTAAGCGAAGATC 81365
Qy 44 rArgGluLeuValSerSerLeuProSerAsn--AspTyr-TripGlyAspThrMetArgLe 63
Db 81364 AAAAATCTAATCTCTTCACTTCTCTTCAGACAAAATTCACCTGGGTAATGTTGTAA 81305
Qy 63 uTyLysGlyPheTrpGlnMetGlyTyrLeuValProGlyLeuMetAlaPheGluAspAs 83
Db 81304 GTACCAAGAGATGTGGTATATCCGCGCTATTTCCAAAGGTGTCTCAATTTCCAGAAAA 81245
Qy 83 nPheValaArgGluThrAspLeuLeuLeuThrLeuProLysAlaGlyThrThrTr 103
Db 81244 CTTTAAGCCTCAAGACACCGGATATCATGTGTCTTCTTCCCTTAATCGCGCACCACTTG 81185
Qy 103 pThrLysAlaLeuThrPheAlaLeuLeuThrArgAspValAsnHisProSer---SerPr 122
Db 81184 GCTCAAGCGCTTACATTCGCACCTGTCGTAGATCG---AAGCACCTTCTCATGATGA 81128
Qy 122 oThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGlu---TyrLe 141
Db 81127 TCATCATCCACTTCTTCTGTATATCCACAGTCTTTTACCCTCCCTTGGATGTAICT 81068
Qy 141 uTyMetGlyArgGluAnthrMetProAspLeuAspMetLeuAnGluSerProArgLe 161
Db 81067 CTAICTATGTAGCGAAAT-----CCCGACCTTACCAGGTCTCATCATCTTCGAGGCT 81014

```

```

Qy 161 uPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLy 181
Db 81013 GTTTTCGACACACATGCGCTCACATACATTGCAAGAGGCTCTCAAGGTTCTACTGTAA 80954
Qy 181 sIleAlaAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLysPhe----- 199
Db 80953 AATTGTGTATATGTCTAGAAACGTAAGACACATTCGTTTTCATATTGGCATTTCTTTTG 80894
Qy 200 -----GlyAsnLeuIleAsnProAspLysLeuLeuAspLeuGluLys 214
Db 80893 TAAGAAACAACCCGATGATACATAATAAGC-----AGTGTGAGGATAC 80849
Qy 214 rValAspIlePheAlaSerGlyLysSerPheCysGlyProGluTrpAsnGlnAlaG 234
Db 80848 ATTTCAGATGTTTGTAGGGAGTCAACTTTTTCGGCCCTTTTGGGACCATGTCCTTAAG 80789
Qy 234 uPheThrAsnAlaAlaSerThrAsnSerAsn---LeuLeuLeuLeuSerTyrGluGluMe 253
Db 80788 CTACTGGAGAGGAAGCTTGGAGATCCAAACCATGCTCTTTATGAGTTTGAGGAGAT 80729
Qy 253 tLeuGluLysProValGluAsnValLysLeuAlaGluPheMetGlyCysGlyPheTh 273
Db 80728 GAAAGAAGAACCTCGTCAGCAGATCAAGAGACTAGCGAATTCCTTAGGATGCTCTTTTAC 80669
Qy 273 rAspAspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPheAspAs 293
Db 80668 TAAGAGAAGAGAAAGCGCTTGGTGGATGAGATTATCATCTTGTCTCTACGTAA 80609
Qy 293 nLeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysLysLeuAspAsnLy 313
Db 80608 TCTGAGCAGTTTGGAGATCAATAAGACCGGAAATTCGATAGTAGTACTGCTAGAGAGACAA 80549
Qy 313 sHisPhePheArgLysGlyGluValaArgAspTTPAlaAsnTyrLeuThrSerGluMetI 333
Db 80548 AACATTTTTCGTAAGGAGAAAGTTGGTACTGGAAGAACTATTGTACTCTCTGAATGGA 80489
Qy 333 eLysLysLeuGluThrAlaGlyLysIleAsnGluSerGluLysHisLeuLeuSerAsnLy 353
Db 80488 GAACAAAATCGAC-----ATGATCATTTCAAGA 80462
Qy 353 sAsnLeuThrAsnPheAsnLeuAsnAsn---TyrCysGluLysSerAsnIleSerLeu-C 372
Db 80461 GAAACTTCAAAACTCTGTTTGAATTCCTGAGTTGTGTGCTTTGTGAGCTTTCTTTAAT 80402
Qy 372 yLeu-----LeuSerTyrSerPheValIleAsnAsnPheIleLeuLeuLysL 388
Db 80401 GCAITGTAATAATAAGACTACTTTTATGTTGTTATATTCCTTCTGTATGCTCTCGT 80342
Qy 388 yLeuLys 390
Db 80341 GTAAGCGT 80334

RESULT 5
AC144726/c
LOCUS AC144726.6 GI:34365847
DEFINITION Medicago truncatula clone mth2-7k13, complete sequence.
ACCESSION AC144726
VERSION AC144726.6
KEYWORDS HTG.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 138586)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
TITLE Medicago truncatula BAC Clone mth2-7k13
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 138586)
AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,

```

```

Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (14-MAY-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS
3 (bases 1 to 138586)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (15-JUL-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
REFERENCE
AUTHORS
4 (bases 1 to 138586)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (30-AUG-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Aug 30, 2003 this sequence version replaced gi:32567849.
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
Location/Qualifiers
1..138586
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-7k13"
/clone_lib="Medicago truncatula BAC library H2"
ORIGIN

Alignment Scores:
Pred. No.: 4,74e-39 Length: 138586
Score: 594.00 Matches: 138
Percent Similarity: 58.79% Conservative: 76
Best Local Similarity: 37.91% Mismatches: 133
Query Match: 28.85% Indels: 17
DB: 9 Gaps: 9

US-09-854-122-16 (1-391) x AC144726 (1-138586)
QY 34 LysGluGluAspSerLysMetTyrLysArgTyrArgGluLeuValSerSerLeuProSer 53
Db 33772 AAAGAAGATGAC-----TTAGGGCAAGATGCAAGGACTTGATACAAACCTTGCATTA 33719
QY 54 AsnAspTyrTrpGlyAsp---ThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyr 72
Db 33718 GTAGAGGGTGGATCGATCCACCTTTTCATGAATATCAAGGATTTGGTTTCTCCCAAG 33659
QY 73 LeuValProGlyLeuMetAlaPheGluAspAsnPheLysAlaArgGluThrAspIleIle 92
Db 33658 ATTCTCAAGGTGCTTTATCTTTTCAAAAGACCTTTCAAGCCATTGATACAGATATTATC 33599
QY 93 LeuThrThrLeuProLysAlaGlyThrTrpThrLysAlaLeuThrPheAlaIleLeu 112
Db 33598 TTATGTACAAATCCCAATCAGTACCACTGCTGCTCAAGCAATGATCTTTGCTTGATA 33539
QY 113 ThrArgAspValAsnHisProSerSerProThrHisProLeuLeuPhePheAsnProHis 132
Db 33538 AACCGTAACAATATCCAAATATTATCAAAAAATCATCTTTTACTCATCTCCACCTCAT 33479
QY 133 SerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeu 152
Db 33478 GTTCTTGTCTTTCATGGAGATTAACTTTACTATGAACTGATATTGTTCTCTGAATC 33419
QY 153 AspMetLeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuPro 172
Db 33418 AATTCACTATCC---CTTCAAGAATCTTCTCCACACACATCTTATGTTTGTGGCA 33362
QY 173 AlaSerValLeuLysSerGlyThrLysIleIleAsnIleSerArgAsnArgLysSerThr 192

```

```

Db 33361 AAATCTTTGAAAGAGTCCAACTGTAAGTGTGTATCTTTGTAGAGATCCCTAAAGACACT 33302
QY 193 PheValSerPheTrpLysPheGlyAsnLeuIleAsnProAspLys-----LeuLeuAsp 210
Db 33301 TTGTGTTTCACTTGGCATTTCTCAACCAAGCTAAGAGACAAAGTAGAGGACACTTCCA 33242
QY 211 LeuGluLysSerValAspIlePheAlaSerGlyLysPheCysGlyProGluTrpAsn 230
Db 33241 TTAGAAGAGCGCTTTTGAGAGTTTGTAGAGAACACTTCTTATGAGACCGTTTGGGAT 33182
QY 231 PheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSer 249
Db 33181 CATGTATTAGTACTCTGGAAGAAAGCTTGAAGACCGGAGAGGTTATGTTTGTAA 33122
QY 250 TyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGly 269
Db 33121 TATGAAGAATGAAATGAAACCTAATTTGTTTGAAGAGAGATTGCTAAGTTTGGGG 33062
QY 270 CysGlyPheThrAspAspGluLysGlnGlyIleValAspGluIleValLysLeuCys 289
Db 33061 TGTCCATTCTCCGAAGAAGAAATCCAAATGTTGTTGATGATATATTACATTGTGT 33002
QY 290 SerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyrAsnSerLys 309
Db 33001 AGTTTGTAGAGCTTAGCAACTTGGAGATCAATAAGAACGGG---AAAAATTTCTGCGAA 32945
QY 310 IleAspAsnLysHisPhePheArgLysGlyGluValArgAspTrpAlaAsnTyrLeuThr 329
Db 32944 ATTGAACAACAAGAACTTTTCCGCTCTGGCAAGTTGGAGATTGGAANAATCTTCCACA 32885
QY 330 SerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGluSerGluLysHisLeu 349
Db 32884 ACTAAATGTTGAGCAGTATAACATAGTTGCGAA---GAGAAATTTGCTAAACATGGG 32828
QY 350 LeuSer-----AsnLysAsnLeuThrAsnPheAsn---LeuAsnAsnTyr----- 363
Db 32827 TTGAGTTTTGTAGATTTTGAAATCTCTGTTCACTTTATGGCTGATGGCTTTCTTAAT 32768
QY 364 ---CysGluLysSerAsnIleSerLeuCysLeuLeuSerTyrSerPheValIleAsnAs 382
Db 32767 CTTTGTACCTTTCTTTTTCAGCTTTCTGTTGGGAATCAATTTGTTCTTCTGGAT 32708
QY 382 nPheIleLeu 385
Db 32707 ATTCTTTTG 32698

RESULT 6
AC146862 166490 bp DNA linear HTG 05-DEC-2003
LOCUS Medicago truncatula clone mth2-174p12, WORKING DRAFT SEQUENCE, 26
DEFINITION unordered pieces.
ACCESSION AC146862
VERSION AC146862.8 GI:38708066
KEYWORDS HTG: HTGS PHASE1; HTGS DRAFT.
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 166490)
REFERENCE
AUTHORS Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
and Roe,B.A.
Medicago truncatula BAC Clone mth2-174p12
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 166490)
AUTHORS Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
and Roe,B.A.
Direct Submission
Submitted (15-OCT-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA

```

REFERENCE 3 (bases 1 to 166490)  
AUTHORS Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.  
and Roe,B.A.  
TITLE Direct Submission  
JOURNAL Submitted (05-DEC-2003) Department of Chemistry And Biochemistry,  
The University Of Oklahoma, 620 Farrington Oval, Room 208, Norman,  
OK 73019, USA

## COMMENT

On Dec 5, 2003 this sequence version replaced gi:38678574.

----- Genome Center  
Center: Department of Chemistry And Biochemistry  
The University Of Oklahoma  
Center code:UOKNOR

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 26 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

\* 1 2098: contig of 2098 bp in length  
\* 2099 2198: gap of unknown length  
\* 2199 5183: contig of 2985 bp in length  
\* 5183 5283: gap of unknown length  
\* 5283 7288: contig of 2005 bp in length  
\* 7288 7388: gap of unknown length  
\* 7388 9418: contig of 2030 bp in length  
\* 9418 9518: gap of unknown length  
\* 9518 11543: contig of 2025 bp in length  
\* 11543 11643: gap of unknown length  
\* 11643 14312: contig of 2669 bp in length  
\* 14312 17259: contig of 2847 bp in length  
\* 17259 19862: contig of 2503 bp in length  
\* 19862 19963: gap of unknown length  
\* 19963 22492: contig of 2530 bp in length  
\* 22492 22593: gap of unknown length  
\* 22593 25049: contig of 2456 bp in length  
\* 25049 27971: contig of 2822 bp in length  
\* 27971 30512: contig of 2442 bp in length  
\* 30512 30612: gap of unknown length  
\* 30612 33341: contig of 2729 bp in length  
\* 33341 33442: gap of unknown length  
\* 33442 35670: contig of 2229 bp in length  
\* 35670 40178: contig of 4408 bp in length  
\* 40178 42782: contig of 2404 bp in length  
\* 42782 45170: contig of 2388 bp in length  
\* 45170 45270: gap of unknown length  
\* 45270 48759: contig of 3488 bp in length  
\* 48759 52749: contig of 3891 bp in length  
\* 52749 52850: gap of unknown length  
\* 52850 62432: contig of 9583 bp in length  
\* 62432 70569: contig of 8037 bp in length  
\* 70569 70670: gap of unknown length  
\* 70670 86794: contig of 16125 bp in length  
\* 86794 100829: gap of unknown length  
\* 100829 100930: contig of 13935 bp in length  
\* 100930 115789: contig of 14859 bp in length  
\* 115789 143040: gap of unknown length  
\* 143040 143141: contig of 27152 bp in length  
\* 143141 166490: contig of 23350 bp in length.

Location/Qualifiers

source 1. 166490  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:3880"  
/clone="mth2-174p12"  
/clone\_lib="Medicago truncatula BAC library H2"

## ORIGIN

Alignment Scores: 7.14e-39 Length: 166490  
Pred. No.: 593.00 Matches: 133  
Score: 56.06% Conservative: 66  
Percent Similarity: 37.46% Mismatches: 142  
Best Local Similarity: 28.80% Indels: 14  
Query Match: 2 Gaps: 8  
DB: 2  
US-09-854-122-16 (1-391) x AC146862 (1-166490)  
Qy 9 ValAlaCysLeuProLeuIleMetAlaGlyIleLeuAlaLeuGluLysCysPheGlySer 28  
Db 55772 GTGTGGTCTCTCCATGCTTCAATGGCTTCAAGATCTCACAAGATTAAGAGACCAATCA 55831  
Qy 29 LysAsnGluGlnGluLys---GluGluAspSerLysMetTyrLysArgTyrArgGluIle 47  
Db 55832 AGGATGACACAGAAACAAATTAGCGAAGAAACAGAGTAATCCATGATTATAACCTA 55891  
Qy 48 ValSerSerLeuPro---SerAsnAspTyrTyrGlyAspThrMetArgLeuTyrLysGly 66  
Db 55892 ATCTGTGCTCTCTCTAGAGAAATGGATGTGATCTCAATATATTTATTTATCATGGA 55951  
Qy 67 PheTyrGlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPhelysAla 86  
Db 55952 TTTTGGTGCCATCACTCTTATCCAACTCTTGAACTCTTTCAAAATAATTTTCATGCC 56011  
Qy 87 ArgGluThrAspIleLeuThrLeuProLysAlaGlyThrThrTyrThrLysAla 106  
Db 56012 AAAGATAGTATATTGTGATGGTTCATCCATCCAAATCTGGCACTACTTGGCTTAAGGT 56071  
Qy 107 LeuThrPheAlaIleLeuThrArgAspValAsnHisProSerSerProThrHisProLeu 126  
Db 56072 CTTCATATGCTATTGTTAATCGCAACACTTTTACCTCCTTAGAAGATAACCATCCATTA 56131  
Qy 127 LeuPhePheAsnProHisSerCysValGluAsnLeuGluTyrLeuTyrMetGlyArgGlu 146  
Db 56132 CTATCATTTAATCCACATGAACCTTGTCCTCATTTTGAAGTAAGTGAATAATACAGAT 56191  
Qy 147 AsnThrMetProAspLeuAspMetLeuAsn---GluSerProArgLeuPheAlaGlyHis 165  
Db 56192 GGTCAATGCTCAATGTATGATGACCAAAATATGGTTCAACCAAGCTTTTATGGAACCTAC 56251  
Qy 166 IleProTyrSerLeuProAlaSerValLeuLysSerGlyThrLysIleIleAsnIle 185  
Db 56252 ATGCCAATTTCTCGTGGCCCACTCAATTCAGAGATCCAATTTGAAGAGTACTCTGAAGGGA 56311  
Qy 186 SerArgAsnArgLysSerThrPheValSerPheTyrLysPheGlyAsnLeuIleAsnPro 205  
Db 56312 TGTAGAACCCTATTGTATACATTTGTATCATATTTGGATTTTCATCAACAACTCAGATTA 56371  
Qy 206 AspLysLeu-----LeuAspLeuGluLysSerValAspIlePheAlaSerGlyIle 222  
Db 56372 AAGAAATCTTTAAGTGAATTAACCTTTGGAGGAATCTTTTGAAGAGTACTCTGAAGGGAATA 56431  
Qy 223 SerPheCysGlyProGluTyr---AsnPheGlnAlaGluPheThrAsnAlaAlaSerThr 241  
Db 56432 TGTATTTTGTCTCGTCTGGGCAATATATGTTGGGTACTTGAAGGAGACGATAGAAAGA 56491  
Qy 242 AsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnVal 261  
Db 56492 CCAGATAGGGTTTTTATCTTTGAAATATGAAGACCTTAAGAGAGATGTTAATTTTTCACACG 56551  
Qy 262 LysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGlnGlyIle 281  
Db 56552 AAAAGAAATTGAGAAATTTGTGGGATTTCTCTTACTGAAGAGAGGAAATAATATGGAGTG 56611

QY 282 ValaspGluIleVallyLeuCySerPheAspAsnLeuLysGlnValAsnLys 301  
 Db 56612 ATTCAAAATATAATCAAGTTATCAGCTTTGAGATGATGAAGAAATCAAGTCAAAATAAA 56671  
 QY 302 AsnGlySerSerTyraSerLysLysAspAsnLysHisPhePheArgLysGlyVal 321  
 Db 56672 TCTGGAAGTCA--GCACCTAAGATGAGAGAGAGTTTCTTCGAAAGGAGAAATA 56728  
 QY 322 ArgAspTrpAlaSerTyraSerGluMetIleLysLysLeuGluThrAlaGlyLys 341  
 Db 56729 GGGGATGGGTGAATATCTCTCCCTTCGATGATAAATAATTA-----TCCAAA 56779  
 QY 342 IleAsnGlySerGluLysHisLeuSerAsnLysAsnLeuThr 356  
 Db 56780 GTCATGGAAGAAAA-----TTAAGTGGATCAAGCCTATCA 56815

RESULT 7  
 AC119410/c  
 LOCUS AC119410 107440 bp DNA linear HTG 18-DEC-2003  
 DEFINITION Medicago truncatula clone mth2-13n10, WORKING DRAFT SEQUENCE, 8  
 unorderd pieces.  
 ACCESSION AC119410  
 VERSION AC119410.5 GI:40018695  
 KEYWORDS HTG: HTGS PHASE1; HTGS DRAFT.  
 SOURCE Medicago truncatula (barrrel medic)  
 ORGANISM Medicago truncatula  
 Sukayoka; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;  
 Medicago.  
 REFERENCE 1 (bases 1 to 107440)  
 AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,  
 Cook,D., Kim,D. and Roe,B.A.  
 TITLE Medicago truncatula BAC Clone mth2-13n10  
 JOURNAL unpublished  
 REFERENCE 2 (bases 1 to 107440)  
 AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,  
 Cook,D., Kim,D. and Roe,B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (26-APR-2002) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 REFERENCE 3 (bases 1 to 107440)  
 AUTHORS Shaull,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,  
 Cook,D., Kim,D. and Roe,B.A.  
 TITLE Direct Submission  
 JOURNAL Submitted (18-DEC-2003) Department of Chemistry And Biochemistry,  
 The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
 OK 73019, USA  
 COMMENT On Dec 18, 2003 this sequence version replaced gi:30231297.  
 ----- Genome Center  
 Center: Department Of Chemistry And Biochemistry  
 The University Of Oklahoma  
 Center code:UOKNOR

\* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 8 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 6423: contig of 6423 bp in length  
 \* 6424 6523: gap of unknown length  
 \* 6524 14616: contig of 8093 bp in length  
 \* 14617 14716: gap of unknown length  
 \* 14717 22860: contig of 8164 bp in length  
 \* 22861 22980: gap of unknown length  
 \* 22981 38252: contig of 15272 bp in length  
 \* 38253 38352: gap of unknown length  
 \* 38353 54847: contig of 16495 bp in length

## FEATURES

source

1..107440

/organism="Medicago truncatula"

/mol\_type="genomic DNA"

/db\_xref="taxon:3880"

/clone="mth2-13n10"

/clone\_lib="Medicago truncatula SAC library H2"

## ORIGIN

Alignment Scores:

Pred. No.: 1 4e-38 Length: 107440

Score: 587.00 Matches: 128

Percent Similarity: 57.14% Conservative: 64

Best Local Similarity: 38.10% Mismatches: 130

Query Match: 28.51% Indels: 28 14

DB: 2 Gaps: 8

US-09-854-122-16 (1-391) x AC119410 (1-107440)

QY 28 SerLysAsnGluGlnGlyLys---GluGluAspSerLysMetTyLysArgTyArgGlu 46  
 Db 22827 TCAAGGGATGGACAGAACCAATTAGCGAAGAAACAGGTAATCCATGATTAATAAAC 22768  
 QY 47 IleValSerSerLeuPro---SerAsnAspTyTrpGlyAspThrMetArgLeuTyLys 65  
 Db 22767 CTAATCTTGCTCTCTCTCTAGAGAAATGGATGATCACTCAATATTTTATCAT 22708  
 QY 66 GlyPheTrpGlnMetGlyTyLeuValProGlyIleMetAlaPheGluAspAsnPheLys 85  
 Db 22707 GGATTTGGTGCCCATCACTCTTATCCATCTGTGAACCTCTTTCAAAATAATTTTAT 22648  
 QY 86 AlaArgGluThrAspIleLeuThrThrLeuProLysAlaGlyThrThrTrpThrLys 105  
 Db 22647 GCCAAAGATAGATATTGTTGTCATCCATGCAATGCTGCAAAATCTGCACACTACTTGGCTTAAA 22588  
 QY 106 AlaLeuThrPheAlaIleLeuThrArgAspValAsnHisProSerSerProThrHisPro 125  
 Db 22587 GGTCTTGCAATGCTATTGTTAATGCCCAACATTTTACCTCCTTAGAGATAACCTCCA 22528  
 QY 126 LeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyLeuTyMetGlyArg 145  
 Db 22527 TTACTATCATTTTATCCACATGAACTTGTGCTCATTTTGAAGTAAGTGGTAATAATACA 22468  
 QY 146 GluAsnThrMetProAspLeuAspMetLeuAsn---GluSerProArgLeuPheAlaGly 164  
 Db 22467 GATGTCAATAGCCTCAAAATGATGTACCAATATGGTTGACCAAGGCCTTTATGGAACT 22408  
 QY 165 HisIleProTySerLeuLeuProAlaSerValLeuLysSerGlyThrLysIleIleAsn 184  
 Db 22407 CACATGCCATTTCTTCGTTGCCCAAGTCAATTCAGAGTCCAATTTGAAATAATTTAT 22348  
 QY 185 IleSerArgAsnArgLysSerThrPheValSerPheTrpLysPheGlyAsnLeuLeuAsn 204  
 Db 22347 ATATGTAGAAACCCATTGATACATTTGTATCATTTTGGACTTTTCATCAACAACCTCAGA 22288  
 QY 205 ProAspLysLeu-----LeuAspLeuGluLysSerValAspIlePheAlaSerGly 221  
 Db 22287 TTAAGAAATCTTTAACTGAAATTAACCTTGGAGGAATCTTTGAAAGGTACTGTAGGGA 22228  
 QY 222 IleSerPheCysGlyProGluTrp---AsnPheGlnAlaGluPheThrAsnAlaAlaSer 240  
 Db 22227 ATATGTTTTTTGGTCCCTTCTGGACATATATGTTGGTACTTGAAGGAGAGCATAGAA 22168  
 QY 241 ThrAsnSerAsnLeuLeuLeuSerTyrgluGluMetLeuGluLysProValGluAsn 260  
 Db 22167 AGACCAGATAGGGTTTTTATTCTTGAAATATGAAACCTTAAAGAGATGTTATTTTTCAC 22108

```

QY 261 ValLysLeuLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGlySerGlnGly 280
Db 22107 ACAGAAAGATTCAGAAATTTGGGATTCCTTTCATCTGAGAGAGAGAAATAATGGA 22048
QY 281 IleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnValAsn 300
Db 22047 GTGATTGAAATATATCAATCAAGTTATGCAGCTTTGAGAGTATGAGGAATCAAGTGAAAT 21988
QY 301 LysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHisPhePheArgLysGlyGlu 320
Db 21987 AAATCGGAACATGCA---GCATTAAGATTGAGAGAGATTTTCTTCGAAAGGAGAA 21931
QY 321 ValArgAspTyrAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGluThrAlaGly 340
Db 21930 ATAGGGGATGGGTGAATATCTCTCCCTTCGATGATAAAAAATTA-----TCC 21880
QY 341 LysIleAsnGluSerGluLysHisLeuSerLysAsnLeuThr 356
Db 21879 AAAGTCATGGAAGAAAA-----TTAAGTGATCAAGCCTATCA 21841

RESULT 8
AC144726 138586 bp DNA linear PLN 30-AUG-2003
LOCUS Medicago truncatula clone mth2-7k13, complete sequence.
AC144726
AC144726.6 GI:34365847
HTG.
KEYWORDS Medicago truncatula (barrel medic)
SOURCE Medicago truncatula
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliopsida; eudicotyledons; core eudicots;
Rosids; eurosids 1; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE
1 (bases 1 to 138586)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Medicago truncatula BAC Clone mth2-7k13
Unpublished
2 (bases 1 to 138586)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (15-JUL-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
3 (bases 1 to 138586)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (15-JUL-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
4 (bases 1 to 138586)
Shauli,S., Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B.,
Cook,D., Kim,D. and Roe,B.A.
Direct Submission
Submitted (30-AUG-2003) Department Of Chemistry And Biochemistry,
The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
On Aug 30, 2003 this sequence version replaced gi:32567849.
----- Genome Center
Center: Department Of Chemistry And Biochemistry
The University Of Oklahoma
Center code:UOKNOR
-----
FEATURES
Source
1. 138586
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/db_xref="taxon:3880"
/clone="mth2-7k13"
/clone_lib="Medicago truncatula BAC library H2"

ORIGIN

```

```

Alignment Scores:
Pred. No.: 9,04e-38 Length: 138586
Score: 579.00 Matches: 138
Percent Similarity: 58.33% Conservative: 72
Best Local Similarity: 38.33% Mismatches: 131
Query Match: 28.12% Indels: 20
DB: Gaps: 9

US-09-854-122-16 (1-391) x AC144726 (1-138586)
QY 36 GluAspSerLysMetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAsp 55
Db 38384 CAAGATGATGACATTAGCCCAAGAAATGCAAGAGATTGATACCCACCTGCCATCAGAAAA 38443
QY 56 TyrTrpGlyAspThrMetArgLeu---TyrLysGlyPheThrPglmMetGlyTyrLeuVal 74
Db 38444 GCCTGGATTTTAAACCCACACCTATCAATCAAGGATTTGGATTGCGACAAAAAATTT 38503
QY 75 ProGlyIleMetAlaPheGluAspAsnPheLysAlaArgGluThrAspIleLeuThr 94
Db 38504 CAAGGTGTTTATCTTGTCTAGAAGCACCTTTCAAGCCCTAGATTTCAGATATCATCTAGTT 38563
QY 95 ThrLeuProLysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArg 114
Db 38564 ACCGTCCTCCAAATCAGGTACCACTTGGCTCAAGCATTGACTTTGCTTTGCTAAACCGC 38623
QY 115 AspValAsnHisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCys 134
Db 38624 AACAAATATCCAAATTTATTCATAACACCATCTTGTCTCATTTTCCAATCCCATCTCTTT 38683
QY 135 ValGlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMet 154
Db 38684 GTTCCCTTCTTGAGATTGACCTTTACAGTGATAAGACTATGTTCTCTCACTTAACCTCT 38743
QY 155 LeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSer 174
Db 38744 TTATCTCAT---CCAGACTCTTCTCCACACATATCTCTGTGCTTGTGTTGTTGCCAAATCT 38800
QY 175 ValLeuLysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheVal 194
Db 38801 GTGAAGATCAACTTCAGAGTTGTGTATCTGTGTAGAGATCCTTAAGACATGTTTGT 38860
QY 195 SerPhe-----TyrLysPheGlyAsnLeuLeuAsnProAspLysLeu-----LeuAsp 210
Db 38861 TCACT-TTACTTTGGCATTTTCAAAACAAGTAAAGACACAGATAGACACCACTTCCA 38919
QY 211 LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrp--- 229
Db 38920 TTAGAAGAGGCTTTTGAGAGTTTGTAGAGGAGTGAGTCTTATGAGCCCTTTTGGAG 38979
QY 230 AsnPheGlnAlaGluPheThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSer 249
Db 38980 CATGTATTAGGATATTGGAAGGAAAACCTTGAAAGGCCAGAGAGGTTGATGTTTATAAAA 39039
QY 250 TyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGly 269
Db 39040 TTTGAGAAGATGAAATGAAACCTAGTTTATTATTGAAAAGATGCTGAGTTTCTGGGT 39099
QY 270 CysGlyPheThrAspAspGluGluLysGlnGlyIleValAspGluIleValLysLeuCys 289
Db 39100 TGTCCATCTTTATGAGAGAGATCCAGGCGCATGATCGATGATATATAAATTGTGT 39159
QY 290 SerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyrAsnSerLys 309
Db 39160 AGTTTGAAGACTTAAAGCAACTTGGAGGTCAACAGACC---GTAAGAGTAGTCAACAGGT 39216
QY 310 IleAspAsnLysHisPhePheArgLysGlyGluValArgAspTyrAlaAsnTyrLeuThr 329
Db 39217 GTTGAAGATTAAGCTTTTTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTC 39276
QY 330 SerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGluSerGluLysHis--- 348

```

Db 39277 GCTGAATGATTAAGCAATTAACACTATACCGRA---AATAATTGGTAAACATGGG 39333  
 QY 349 -----LeuSerSerAsnLysAsnLeuThrAsnPhaAsnLeu 360  
 Db 39334 TTGAGTTTGTAGATTTTGGAGTGTGTTTCAAGTCTAAGAAATACATTCGGAGATTA 39393  
 QY 361 AsnAsnTyrCysGluLysSer---AsnIleSerLeuCysLeuLeuSerTyrSerPheVal 379  
 Db 39394 AGTTTTCAGTGAGAGAACTTCACTTATTTTAAAGCTTCTTCTTCCTTCATTCATC 39453

RESULT 9  
 LOCUS AC0271134 102183 bp DNA linear PLN 15-AUG-2000  
 DEFINITION Arabidopsis thaliana chromosome I BAC F13B4 genomic sequence,  
 complete sequence.  
 AC0271134 GI:8576187  
 VERSION AC0271134.4  
 KEYWORDS HTG.  
 SOURCE Arabidopsis thaliana (thale cress)  
 ORGANISM Arabidopsis thaliana  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
 rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
 1 (bases 1 to 102183)  
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,  
 Altati, H., Nguyen, M., Lam, B., Southwick, A., Miranda, M., Brooks, S.,  
 Buehler, E., Chao, Q., Chin, C., Chio, J., Choi, E., Gonzalez, A.,  
 Hwang, B., Johnson-Hopson, C., Khan, S., Kim, C., Koo, T., Lee, J.M.,  
 Lenz, C., Liu, A., Liu, S., Mukharsky, N., Pham, P., Sakano, H.,  
 Shinn, P., Toriumi, M., Vaysberg, M., Walker, M., Yu, G., Ecker, J.,  
 and Davis, R.W.  
 2 (bases 1 to 102183)  
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,  
 Altati, H., Nguyen, M., Lam, B., Southwick, A., Bei, Q., Buehler, E.,  
 Chin, C., Chio, J., Choi, E., Dunn, P., Gonzalez, A., Hwang, B., Kim, C.,  
 Koo, T., Lee, J.M., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S.,  
 Mukharsky, N., Pham, P., Sakano, H., Schwartz, J., Shinn, P.,  
 Thavari, A., Toriumi, M., Vaysberg, M., Walker, M., Yu, G., Ecker, J.,  
 Theologis, A., and Davis, R.W.  
 Direct Submission  
 Submitted (28-MAR-2000) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA  
 3 (bases 1 to 102183)  
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,  
 Altati, H., Nguyen, M., Lam, B., Southwick, A., Bei, Q., Buehler, E.,  
 Chin, C., Chio, J., Choi, E., Dunn, P., Gonzalez, A., Hwang, B., Kim, C.,  
 Koo, T., Lee, J.M., Lenz, C., Li, J., Liu, A., Liu, K., Liu, S.,  
 Mukharsky, N., Pham, P., Sakano, H., Schwartz, J., Shinn, P.,  
 Thavari, A., Toriumi, M., Vaysberg, M., Walker, M., Yu, G., Ecker, J.,  
 Theologis, A., and Davis, R.W.  
 Direct Submission  
 Submitted (21-JUN-2000) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA  
 4 (bases 1 to 102183)  
 Authors Federspiel, N.A., Palm, C.J., Conway, A.B., Conn, L., Hansen, N.F.,  
 Altati, H., Nguyen, M., Lam, B., Southwick, A., Ecker, J., Theologis, A.,  
 and Davis, R.W.  
 Direct Submission  
 Submitted (15-AUG-2000) DNA Sequencing and Technology Center,  
 Stanford University, 855 California Avenue, Palo Alto, CA 94304,  
 USA  
 COMMENT  
 On Jun 21, 2000 this sequence version replaced gi:8134868.  
 Bases 1-55,850 of F13B4 overlap with bases 47,308-103,157 of BAC  
 clone T6J4, gb|AC011810 and bases 76,686-102,183 of F13B4 overlap  
 with bases 1-25,498 of BAC clone F12P23, gb|AC027656.  
 e-mail for correspondence: arabseq@sequence.stanford.edu  
 Genes with  
 similarity to proteins in the databases are named 'putative',  
 'like' or 'similar to'. Genes that have EST similarity but no  
 significant protein similarity are described as 'unknown proteins'.  
 Genes that are annotated based only on gene prediction software

are described as 'hypothetical proteins'. The gene prediction  
 programs used to predict genes include: Grail (Informatics Group,  
 Oak Ridge National Laboratory,  
 http://compro.ornl.gov/section/index.html), GENSCAN (Chris Burge,  
 http://genes.mit.edu/GENSCAN.), FEXA (Victor Solovyev,  
 http://genomic.sanger.ac.uk/gf/gf.shtm), and NetPlantGene (S.M.  
 Hebsgaard, et al., CBS, Technical University of  
 Denmark, http://www.cbs.dtu.dk/NetPlantGene.html).

FEATURES  
 Location/Qualifiers  
 1..102183  
 /organism="Arabidopsis thaliana"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3702"  
 /chromosome="I"  
 /clone="F13B4"  
 1..55850  
 /note="overlap with bases 47,308-103,157 of BAC clone  
 T6J4, gb|AC011810. See GenBank record for clone T6J4 for  
 annotation in this region."  
 complement(55518..56867)  
 /gene="F13B4.1"  
 complement(join(55518..56358,56467..56566,56660..56867))  
 /gene="F13B4.1"  
 /note="Hypothetical protein"  
 /codon\_start=1  
 /protein\_id="AAF99819.1"  
 /db\_xref="GI:9802750"  
 /translation="MYVTRKLSSEYORNPSELTPTEGNSGVLIQDESRTPCCFG  
 KCFDKHLGLFPQNTKTVRFQGVVYLDPAFIPVLKPLSSNLYIVIRSGKRT  
 GEACVATGDKTCCICFPIDPDDPDIYQOFEHQRGSSNQFFATVSFAS  
 EGIPFQFLTRKGVAAISLPEEYGLIDDAKGVDAKLVELPDLDSVWVKRWYV  
 FVKEGDANDQMKSMYSMTLQRPFEVFCNIGNNKHFVVDVEVEVVKLEG  
 DKIAETKGVNSDGVVFSVSGTOKIGLGSVLERMKWEERFGWSKDLRSIKRT  
 BKFEGGGPHKSYCYVLTVELFKRTDGLSLVTEPKHVDKLKSKWD"  
 complement(56592..61490)  
 /gene="F13B4.2"  
 complement(join(58592..59438,59566..59665,59759..59905,  
 60390..61084,61171..61490))  
 /gene="F13B4.2"  
 /note="Hypothetical protein"  
 /codon\_start=1  
 /protein\_id="AAF99820.1"  
 /db\_xref="GI:9802751"  
 /translation="MYVTRKLSSEYORNPSELTPTEGNSGVLIQDESRTPCCFG  
 GLCLVDSYLSGLPELQNLKLAARFAGEDDADCDPVFIPVLKPLSSNLYAIRV  
 GAGGAPASGKSDRVPCCSCSYVPEAKQDLPYDMYQFMHOKKSSRYIAT  
 SVAPGVSPFLKXNKTVEYSRFDPELKDAGLNTBELTLPGLGMSVWVKWYV  
 PFIPYKRDVEDQIKRSMYISMTLEQWEVFSYEDKSRNYDLVDVLDTEVVKYD  
 GQISRGVGEAFVAVGDKKIGLGSVVERMKWEERFGWTKGDNERSVAKRLE  
 XSTDGPNSGVLIQDESRTPKCCFVNSFDMELNGLPFPQNTKTVIRKQIGTALYL  
 DEVARIPVLHOPPSNLYIVIRSGKTHGACVSAKEDRASCCFCFTIPEAVRPL  
 DFDLQOQFEHQSRSTKFPATSDGTPQPLTRKGSVTFESAPEVDGLIDD  
 AKGIVDAKLRIEPLDLSVSVVGRWYVFLFVKEGDANDQMKSMYSMTLQRPFEV  
 FFCENGNKRFVVDVEVEVTEVVKLEGERIARETGQVNSDGVVWFNAGNEKIGLSI  
 SYVLERMKWEERFGWLNKGYEQRSIKRTFERFEGGPGYWKSYRCYVLIQDESRTP  
 GSLVLYTEFKHVDKLKSKWD"  
 complement(62660..65932)  
 /gene="F13B4.3"  
 complement(join(62660..63503,63707..64040,64557..65380,  
 65510..65621,65725..65932))  
 /gene="F13B4.3"  
 /note="Hypothetical protein"  
 /codon\_start=1  
 /protein\_id="AAF99821.1"  
 /db\_xref="GI:9802752"  
 /translation="MYVTRKLSSEYORNPSELTPTEGNSGVLIQDESRTPCCFG  
 KCFGDNLGLFPQNTKTVIRKQIGHGDRIVLDSVAFIPVLHQPSSNLYVIRRR  
 GHGACVSAKEDKSCCICTFVSNAAPRLDLPDIYQOFEHQRGSSNSTRKFPAT  
 SVASGCIPLRFRKQWTPPSASQDGLIDDAKGVDAKLVELPDLDSVWVKWYV  
 VFLFVKGDANDQMKSMYSMTLHORFEVFFCENVDKNKEVIVDVEVEVVKL  
 EGENIGTKGVNSDGVVFGVSGTQKIGISVLERMKWEERFGWSKXGNEKJLSI  
 KTEFEGGGQWKSRYCYVLTVELFKRTDGLSLVTEFKHVDKLKSKWYV  
 LSEYQNSDLSLTPTEGNSGVLIQDESRTPCCFGNCVYDGLKLPFQNAKLTV



TYRTGTGNDRSYHDQVLPFVLDQPIYSNRYVYIKRKGKSGGASAKKEEDRVPC  
 FCFSYVPEKQADPYDIYQOFVHORPKSSGYDTATSVAPNGIPFELFKRKWTVE  
 YSNQDGLITDAKINTLRSLKPNDMNTSVVGVKMYVPEFVKERDAKQIKSSY  
 YSMILQKWEYYSVUNVYQKGEVVDVDTQVVKLEGQETISRETSQGGFWWF  
 SVLGERODKIKGLGSVVRMKBEERFVGLNNGERSNIKRSEFEGSSHMKSRYC  
 LVLIJESFLKMDGSLVLTVEFTHYDKLKSWD  
 complement (66277..69118)  
 /gene="F13B4.4"  
 complement (join (66277..67099,67189..67554,67689..67786,  
 67968..68724,68802..69118))  
 /gene="F13B4.4"  
 /note="Unknown protein"  
 /codon\_start=1  
 /protein\_id="AAF99822.1"  
 /db\_xref="GI:9802753"  
 /translation="MIVTRRLSEYKRNPSLKKPPPEGNSGLVVIQDESLITCCFG  
 LCNELFLGLPFPQNAALTRDRDQGNTHVRDPVFIPLVGLPSSNRYTYERCG  
 YYSRESSTSEVERVCCFIRFNHYMKKPPQDPDYHQVYITITSSPRCY  
 FAKVPANALLPAFLMQQETIENSTLQDFGLIDDAKGLNLESLRSLGLDSVVV  
 GWVYVPIFKVGGDIIDVKI SMYVNTLHORWEVFFVYVNHEDCVVVDVDEL  
 VYINLEGOKVSTETTNVDAKGLVFWQVKDKIGLGSVVVERMESEESFGWKVDGNR  
 VVVKLDRREGASSHWKSKVCLVSCSVSEFPMQSGVMIQRGDFVYVLIETDERI  
 KFSKSKTRFSTMTITRSFYRRNPAELTPPEGNSGLVVIQDQHSSTRATSCF  
 GSWLVVDSLSGLPQLNKLAVTNIGDDSDRDPVFIPLVLDKPLSSNRYIAIKRH  
 KHGSEASGAKEDIVSCFCSSFPKAKQKLPDPFMYQOFKHQKPSRSYAT  
 SVAPQVPPFVKKEWTVYSRFQDFELRDAKGLITELRTELLPGLMSTVGKVV  
 PFIYKEREVQKIKNSMYVSMTLQRWKELPSDENKSDNVDVVEVEVEVKLE  
 GOMIRGLFANGFVGGDKKIGLGSVVVERMKBEERFGWTKGNDGNSMTVKRLS  
 KSTDSGLKHYCHYLIESFVLKRMDSLVLTVEFTHYDKLKTWD  
 complement (69433..72830)  
 /gene="F13B4.5"  
 /note="synonym: AAP11"  
 complement (join (69433..69658,69744..69816,69923..69975,  
 70060..70118,70214..70376,70373..70457,70549..70589,  
 70670..70733,70849..70918,71026..71094,71224..71318,  
 71397..71497,71584..71606,71693..71768,71872..71949,  
 72079..72129,72222..72231,72358..72508,72746..72830))  
 /gene="F13B4.5"  
 /product="AAP11"  
 complement (join (69637..69658,69744..69816,69923..69975,  
 70060..70118,70214..70376,70373..70457,70549..70589,  
 70670..70733,70849..70918,71026..71094,71224..71318,  
 71397..71497,71584..71606,71693..71768,71872..71949,  
 72079..72129,72222..72231,72358..72494))  
 /gene="F13B4.5"  
 /notes="Location of cDNA clone gb|AF091843,  
 cholinephosphotransferase; ethanolaminephosphotransferase"  
 /codon\_start=1  
 /product="aminealcoholphosphotransferase"  
 /protein\_id="AAF99823.1"  
 /db\_xref="GI:9802754"  
 /translation="MGYIAGHVAALHRYKYGVDHSYAKYLVQFTRFVKVPLW  
 MPNMTLGMFPLVTSLLGYISQPSDPPRPPVHFVFAHGLLLFLYOTFVADGQA  
 RNTNSSLPGELFDHGDALACAFAMAFSTAMGORDTFNFWI SAIPFYGATWEHY  
 FNTLLIPVINGTEGLALIFVSHFFTAIVGAEWAQOLGQSIPLFSWVPVFNQTS  
 RAVLVNMIAPVITVAENVTVKVRNMGSMVLAAMLYPFVVLGGVLIWDYLS  
 PINLIATYHLVILGTLAFGLVGMILAHLCDEPKLKTNMKSLYLPLPALANAL  
 TAKNAGVFLVDELWLLGVCFTVSLYLHFAITSVHTEALGIYCFRIRKEA"  
 complement (73331..74733)  
 /gene="F13B4.6"  
 complement (join (73331..73621,73703..73861,73933..74733))  
 /gene="F13B4.6"  
 /note="Unknown protein"  
 /codon\_start=1  
 /protein\_id="AAF99824.1"  
 /db\_xref="GI:9802755"

## Alignment Scores:

Pred. No.:	7,01e-38	Length:	102183
Score:	578.50	Matches:	137
Percent Similarity:	56.67%	Conservative:	67
Best Local Similarity:	38.06%	Mismatches:	133
Query Match:	28.10%	Indels:	24

DB: 8 Gaps: 9  
 US-09-854-122-16 (1-391) x AC027134 (1-102183)  
 Qy 31 GluGlnGluLysGluGluAspSerLysMetTyrLysArgTyrArgGluLeuValSerSer 50  
 Db 35707 GAAGAAGAAGAAAGAAAGAAACCAAGTGAAGAAACCAAAAGTTGTGATCTCTTCA 35766  
 Qy 51 LeuProSerAsn--AspTyrTrpGlyAspThrMetArgLysGlyPheTrpGln 69  
 Db 35767 CTTCCCTTCAGACATAGATTGCTCAGGACCAAGTTGTACAGATCAAGGATTTGGTAC 35826  
 Qy 70 MetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPhelysAlaArgGluThr 89  
 Db 35827 GACAAAGATATTCTCCAAGCAATCTCAATTTCAACAAAACATTTTCAGGCCAACGAACG 35886  
 Qy 90 AspileleLeuThrThrLeuProLysAlaGlyIleThrThrThrLysAlaLeuThrPhe 109  
 Db 35887 GATATAATTGCTGCTTTTCCCAATCGGACGACCTGGTCAAGGACATCAATTC 35946  
 Qy 110 AlaIleLeuThrArgAspValAsnHisProSerSerProThrHisProLeuLeuPhePhe 129  
 Db 35947 GCATCGCGCAAGATCAAAACAC-----ACITCAGACAATCATCTCTGTAACTCAT 36000  
 Qy 130 AsnProHisSerCysValGlnAsnLeuGluTyr---LeuTyrMetGlyArgGluAsnThr 148  
 Db 36001 AATCTCATGAGCTAGTGCCTACCTCGAGCTCGATCTTTATCTC-----AAAAGCTCG 36054  
 Qy 149 MetProAspLeuAspMetLeu---AsnGluSerProArgLeuPheAlaGlyHisIlePro 167  
 Db 36055 AAACCGGATTTGACCAAGTTGCCATCATCTCCGAGATTGTTCTCAACCCACATGTC 36114  
 Qy 168 TyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIlelelelelelelelele 187  
 Db 36115 TTTGATCGCTTAAAGTACCGCTTCAAAGATCTCTTTCGAAGATCGTGTACGTGTCAGG 36174  
 Qy 188 AsnArgLysSerThrPheValSerPheThrLysPheGlyAsnLeuLeuAsnProAspLys 207  
 Db 36175 AACGTAAAGAGCTATTGGTATCATCTTGTGTTTCGAAAACCTCCATGATGGAGAAAC 36234  
 Qy 208 LeuLeuAspLeuGluLysSerValAspIlePheAlaSerGlyLysSerPheCysGlyPro 227  
 Db 36235 AATTAAAGTCTCGAGGCTTTGTTTCGAGCTTTTATGTAGCGAGTTAACTTATCGGTC 36294  
 Qy 228 GluTrp---AsnPheGlnAlaGluPheThrAsnAlaAlaSerThrAsnSerAsnLeuLeu 246  
 Db 36295 TTATGGAAATATGTTAGGCTATTGGAGAGGAGCTTGGAGATCCTAAGCATGTCCT 36354  
 Qy 247 LeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGlu 266  
 Db 36355 TTCTTGAGGTACGAGGAGTTGAGACGCGCTCGTGTGCAAAATCAAGAGATTGCAGAG 36414  
 Qy 267 PheMetGlyCysGlyPheThrAspAspGluGluLysGlnGlyIleValAspGluIleVal 286  
 Db 36415 TTCTTAGATTGTCATTCACAAAGAAAGAAAGATAGTAGGAGGTGTAGACAGATCTTG 36474  
 Qy 287 LysLeuCysSerPheAspAsnLeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyr 306  
 Db 36475 GAATTTGTTCTCTAAGAAACCTTAGCGGTTTGGAGATCAACAAACAGCA---AGCTTG 36531  
 Qy 307 AsnSerLysIleAspAsnLysHisPhePheArgLysGlyGluValArgAspTrpAlaAsn 326  
 Db 36532 TCGGAAGGAGTAAGTTTCAAGAGATTTTTCGTAAGGGAAGTGTGTGATTTGGAAGAGT 36591  
 Qy 327 TyrLeuThrSerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGluSerGlu 346  
 Db 36592 TATATGACTCTGAAATGGAAAACAAATCGAC-----ATGATTGTTGAGGAG 36639  
 Qy 347 LysHisLeuLeuSerAsnLysAsnLeuThrAsnPheAsnLeuAsnAsnTyrCysGluLys 366  
 Db 36640 AAACCTTCAAGGCTCTGGTTT-GAAATGTAGAGTTC-----GAGGAGGAGGAG 36674  
 Qy 367 SerAsnIleSerLeuCysLeuLeuSerTyrSerPheValIleAsnAsnPheIleLeuLeu 386







complete sequence.  
AC016662  
VERSION AC016662.7 GI:12325130  
KEYWORDS  
SOURCE Arabidopsis thaliana (thale cress)  
ORGANISM Arabidopsis thaliana  
REFERENCE Rukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
1 (bases 1 to 108387)  
AUTHORS Lin, X., Kaul, S., Town, C.D., Benito, M.-I., Creasy, T.H., Haas, B.J., Wad, D., Maiti, R., Roming, C.M., Koo, H., Fujii, C.Y., Uterback, T.R., Barnstead, M.E., Bowman, C.L., White, O., Nierman, W.C. and Fraser, C.M.  
Arabidopsis thaliana chromosome 1 BAC F2P9 genomic sequence  
Unpublished  
2 (bases 1 to 108387)  
TITLE Arabidopsis thaliana chromosome 1  
AUTHORS Lin, X. and Kaul, S.  
TITLE Direct Submission  
JOURNAL Submitted (04-DEC-1999) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, xlin@tigr.org  
REFERENCE 3 (bases 1 to 108387)  
AUTHORS Town, C.D. and Kaul, S.  
TITLE Direct Submission  
JOURNAL Submitted (19-JAN-2001) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA, cdtown@tigr.org  
COMMENT On Jan 19, 2001 this sequence version replaced gi:12280816.  
Address all correspondence to: at@tigr.org

BAC clone F2P9 is from Arabidopsis thaliana chromosome 1  
The orientation of the sequence is from Sp6 to T7 end of the BAC clone.  
Genes were identified by a combination of several methods: Gene prediction programs including Genscan+ (Chris Burge, <http://CCR-081.mit.edu/GENSCAN.html>), GeneMarkMM (Mark Borodovsky, <http://genemark.biology.gatech.edu/GeneMark/>), GlimmerA (a variant of GlimmerM, see Mihaela Pertea,  
<http://www.tigr.org/softlab/glimmerm.htm>), and GeneSplicer (Mihaela Pertea and Steven Salzberg, contact mpertea@tigr.org), searches of the complete sequence against a peptide database and the plant EST database at TIGR (<http://www.tigr.org/tdb/tgi.shtml>). Annotated genes are named to indicate the level of evidence for their annotation. Genes with similarity to other proteins are named after the database hits. Genes without significant peptide similarity but with EST similarity are named as unknown proteins. Genes without protein or EST similarity, that are predicted by more than two gene prediction programs over most of their length are annotated as hypothetical proteins. Genes encoding tRNAs are predicted by tRNAscan-SE (Sean Eddy, <http://genome.wustl.edu/eddy/tRNAscan-SE/>). Simple repeats are identified by RepeatMasker (Arian Smit, <http://ftp.genome.washington.edu/RM/RepeatMasker.html>).  
Location/Qualifiers  
1..108387  
/organism="Arabidopsis thaliana"  
/mol\_type="genomic DNA"  
/cultivar="Columbia"  
/db\_xref="taxon:3702"  
/chromosome="1"  
/clone="F2P9"  
35..1255  
/gene="F2P9.1"  
join(<35..663,669..1255)  
/gene="F2P9.3"  
join(<35..663,669..1255)  
/gene="F2P9.1"  
/codon\_start=2  
/product="unknown protein, 5' partial; 35-1255"  
/protein\_id="AA052507.1"  
/db\_xref="GI:12325131"  
/translation="KRLLANLQRYGFPSSLOHFLSRNHLNLSLVEISLIGIL  
LSLIPKQSLVSLSDPCNVLRSEFLKRWVPLSNCGHGVVSSSAIKSLVHSSRIIG  
IGPKFNECVRLKSGFCDSVRLSSGPGVLYNEIEIRKTEFLVGTGIARDNI  
ERFHPVFEVLGTETRLKPLLDLDFMKNGFSKDDVKKIAERERFLVNSLPRCLIE

INTLCKREIVRSIIISGAPRAGPEVKLRVDCLCKGLIRRDFAKVVWKEPRVILYEI  
EDIEKKIEFLNRMGHINCLADVPEVLGNVLOKQIVPRVNVVDLYLKGGLGCDIGL  
KGLKPSMKREYNLYVNPYECERIFGKRKENVYNKHPAGLWKLKMKPSPNLTTKED  
VQNMKSFIIESLA"  
2203..2222  
/rpt\_family="(A)n"  
4810..6511  
/gene="F2P9.2"  
/note="similar to cytochrome P-450 GB:AA037231 from  
[Phalaenopsis sp. SM9108]"  
join(<4810..5811,5900..6511)  
/gene="F2P9.2"  
join(4810..5811,5900..6511)  
/gene="F2P9.2"  
/codon\_start=1  
/product="putative cytochrome P-450; 4810-6511"  
/protein\_id="AA052514.1"  
/db\_xref="GI:12325138"  
/translation="MTIDMVLSPASRSGSPFFSLELCLSLFISLFVFLVTPGGFA  
WALYKARHTPESTGTQPAIPGSPGLPIFGLLAFVNNALTHRLILANIADTCAKALM  
AFVGSFTFVITSPEETAKELLASSAFADSPKESAVELLFDRAMGPAFGDYREL  
RISSTHLPSPKRIFFSSGESRKKQNNVGEIKNAECYGRVHAKTILHFGSLWMS  
VGRKYNNEGIVIKSKSELEHLVSGEYLLGIFNWSHFFPGRWLDLQGVRRCS  
LVGRVNVFGKIIINDHKKSLRDNPSESTYDDDFVDVLGMHNSKLSLMDLAVLW  
EMIFGTDVAILLEWLARVLPDIOAKAABIDCIVGDSGRQVTSDLPLKPYVR  
AIVKTLRMHPGGLLSWRLSIHDTQIGTHFIPAGTANVMAITHDSKVPFAHE  
YKPERFLGASQNNFPIMGSLRLAPFGAGRRVCPGKMGCLATVLMWLAQLLGSYKVV  
SCGEVDLSETLKLSEMKNTLVCAIPRG"  
complement(6762..6783)  
/rpt\_family="AT-rich"  
7673..8831  
/gene="F2P9.3"  
/note="similar to flavonol 4'-sulfotransferase GB:P52837  
from [Flaveria chloraefolia]"  
7673..8831  
/gene="F2P9.3"  
7673..8689  
/gene="F2P9.3"  
/codon\_start=1  
/product="putative flavonol sulfotransferase; 7673-8689"  
/protein\_id="AA052515.1"  
/db\_xref="GI:12325136"  
/translation="MESKTTQNGSEVVELTEFEKTKQYQDFIATLPKSGWRPDEIL  
TOYQGWQECLEGLHAKHFAEPTDFVLSYPTGTTWTKALTYAIWVNSRYDD  
AANLKNPHEFVYVEIDFAFYTDVLDQRKNPLFSTHINGLDPSIVNSGCKM  
VYMRDQDTFISMTWTLHREKSGSGLASLSEDFMEFKGLSVYGYLHVLCYKKA  
YOENDRILFLRYETMRANPLPFVKLAESMGYGTDEERENGVAEKVKLCSFETLK  
NLEANGKREDRPAVYANSAYFRKGVGDWANYLTPMAARIDGLVEEKFKDTGLL  
QHDN"  
complement(9123..10175)  
/gene="F2P9.4"  
/note="similar to FLAVONOL 4'-SULFOTRANSFERASE GB:P52837  
from [Flaveria chloraefolia]"  
complement(<9123..10175)  
/gene="F2P9.4"  
complement(9123..10175)  
/gene="F2P9.4"  
/codon\_start=1  
/product="putative flavonol sulfotransferase; 10175-9123"  
/protein\_id="AA052515.1"  
/db\_xref="GI:12325139"  
/translation="MESETLTAKATITTTLPSSHDETKTSTPEKQKQYQDLISTF  
PHEKGRPKPELIEVGYWMLPSLEGICHAQEFPOARPSDFLVCYPTGTGTTWTKAL  
THAIANRFRDSSNPLKRNPFYIEIDFPFVVDVLKDKGNTLPSTHPIYEL  
LPDSVSKSGKVIYWRPKDTFTFSMTFLHKEITELGPNVSNLEESFDMRCGLSGYV  
PYLNHILAYKWAYQENPDRIILFLETNRADPLFYKSLARFNGHGFIASEEKGVVR  
KYVNLCSTFETLNLEANKGEKREDRPGVANSAYFRKGVGDWNSYLTPEMAARIDG  
LMEKFKTGTLLEHGK"  
complement(12228..12256)  
/rpt\_family="AT rich"  
complement(12462..12498)  
/rpt\_family="AT rich"  
complement(12549..13136)  
/rpt\_family="AT rich"



RESULT 12  
AY087493

AY087493	1347 bp	mRNA	linear	PLN 14-APR-2003
LOCUS				
DEFINITION	Arabidopsis thaliana clone 36020 mRNA, complete sequence.			

ACCESSION AY087493  
VERSION AY087493.1 GI:21406230  
KEYWORDS FLI CDNA.

SOURCE	ORGANISM	REFERENCE
Arabidopsis thaliana (thale cress)	Arabidopsis thaliana	1 (bases 1 to 1347)
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.	Eukaryota; Viridiplantae	1 (bases 1 to 1347)

REFERENCE	AUTHORS	TITLE
1 (ABSTRACTS IN CO 1347)	Haas, B.J., Volfovsky, N., Town, C.D., Troukhan, M., Alexandrov, N., Feldman, K.A., Flavell, R.B., White, O. and Salzberg, S.L.	Full-length messenger RNA sequences greatly improve genome

**JOURNAL** Genome Biol. 3 (6), RESEARCH0029 (2002)  
annotation  
full-length messenger RNA sequences greatly improve genome  
annotation

22088475 MEDLINE  
12093376 PUBMED

2 (bases 1 to 1347)  
 Brover, V., Troukhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and  
 AUTHORS

AUTHORS	TITLE
BOYER, W.; BLOOMBERG, M.; ALEXANDROV, N.; LU, J.-F.; FAVELLI, A. and FELDMANN, K.	Full-length cDNA from <i>Arabidopsis thaliana</i>

Full-length cDNA from *Aradipoopsis thaliana*  
Unpublished  
JOURNAL  
3 (bases 1 to 1347)  
REFERENCE

REFERENCE  
3 (bases 1 to 134)  
AUTHORS  
Brover, V., Troukhan, M., Alexandrov, N., Lu, Y.-P., Flavell, R. and  
Feldmann, K.

TITLE	JOURNAL	DATE	FROM	TO	REMARKS
Direct Submission					
Submitted (11-MAR-2002)	Ceres	INC	1007 Malibu Canyon Road		

JOURNAL  
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,  
Malibu, CA 90265, USA  
This clone sequence is one of 5 000 Ceres full-length cDNAs made

**COMMENT:** This clone sequence is one of 5,000 Ceres full-length cDNAs made available to TIGR and Genbank. The following quality assessment of this set was done by comparison with known proteins: one percent of this set was done by comparison with known proteins.

this set was done by comparison with known proteins: two percent of the clones are estimated to be 5'-truncated; less than one percent

are 3'-truncated; approximately two percent represent alternative splice variants, including unspliced introns and spliced exons; one

percent may contain premature stop codons; five percent may have frame shifts in a coding region. A sequence is considered to be

5'-truncated if it lacks the translation initiation start (ATG). A sequence is considered to be 3'-truncated if it lacks the

C-terminal end of the encoded protein. Please note that these cDNA sequences are derived from the Ws or LAer ecotypes and therefore

sequences are generated from the 5' and 3' ends of the *ccpA* gene and may contain polymorphisms when compared to sequences from Col-0. Genset carried out the library production and sequencing of the

carried out the library production and sequencing of the full-length clones. Ceres, Inc. carried out the clustering of the 5' sequences selection of clones, and sequence assembly.

```

FEATURES
  source
    Location/Qualifiers
      1. .1347
        sequences, selection of clones, and sequence assembly.

```

```
1. .134/  
source  
/organism="Arabidopsis thaliana"  
/mol_type="mrna"
```

```

/db_xref="taxon:3702"
/mol_type="mrna"
/clone="14020"

```

```
CDS
/clone="38020"
190.1206
/codon_start=1
```

```

/codon_start=1
/product="putative flavonol sulfotransferase"
/protein_id="unacc021"

```

```

/protein_id="AAW65036.1"
/db_xref="GI:21593087"

```

```
/translation= MESKTTQNGSEWELTEFEKTKKQYDFIATLPKSGWRPBEIL
TQYGGHWWQCLLEGLFHAQKHFEARPTDFVCSYPKTTGTTWLKALTYAIVNRSYDD
```

AANPLLKRNPHEFVPEIDFAFYPTVDVLQDRKNPLFSTHINCLLPDSIVNSGCKM  
VYIWRDPKOTFI SMWTFLLHKEKSEQQLASLEDSDFMFKGLSVYGVPYLDHVLGYWKA

YQENPDRILFLRYETMRANPLPFVKRLAEFMWGYGTTDEEBENGVAEKVKLCSFETLK  
NLEANKGDEREDRPAVYANSAYFRKGKVGDMWNYLTPEWAARIDGLLEBEFKDTGLL

ORIGIN

Alignment Scores:

Argument Scores:	
Pred. No.:	1.38e-39
Score:	573.00
Length:	1347
Matches:	142

Score:	575.00	Matches:	172
Percent Similarity:	53.10%	Conservative:	72
Best Local Similarity:	35.24%	Mismatches:	152

best local similarity: 35.24% mismatch: 132

\_\_\_\_\_

Query Match:	27.83%	Andals:	38
DB:	8	Gaps:	12
US-09-854-122-16 (1-391) x AY087493 (1-1347)			
QY	14	LeuileMetAlaGlyIleLeuAlaLeuGluLysCysPheGlySerIysAsnGluGln---	32
DB	154	ATCCTCTTGGACTCCCTTCACTCGATCTCATCTGCAATGGAATCAAGACCAACCCAAAAC	213
QY	33	-----GluLysGluGluAspSerLysMetTyrIysArgTyrArgGluIle	47
DB	214	GGATCCGAGTCGTGCACTCAGAGTTCGAGAAACCAGAGAGAGTATCAAGATTTC	273
QY	48	ValSerSerLeuProSerAsnAspTyrTrp-----GlyAspThrMetArgLeuTyrIys	65
DB	274	ATCGTACACTTCCAAAGACGAAGCGTCGAGACCGAGATCGATCTTACCCAAATACCGT	333
QY	66	GlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPheLys	85
DB	334	GGACACTGGTGGCAAGAATGTCCTCTCGAAGGTCTTTTCACGCTAAAGACCAATTCGA	393
QY	86	AlaArgGluThrAspIleIleLeuThrTyrLeuProLysAlaGlyThrThrTrpThrLys	105
DB	394	GCAGACCAACTGATTTCCTCGTGTAGTACCCCAAAAACCGGTACAACTTGGCTCAA	453
QY	106	AlaLeuThrPheAlaIleLeuThrArgAspValAsnHisProSerSerProThrHisPro	125
DB	454	GCACACTTACGCAATCGTCATCGT-----TCTCGTTACGACGCGCCGCAAAACCA	507
QY	126	LeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArg	145
DB	508	CTCTCTAAACGAAACCCCTCAGAGTTGTCCTTACGTGAGATCGACTTCGCG-----	561
QY	146	GluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArg---LeuPheAlaGly	164
DB	562	-----TTTTACCACCGTTGATGTCTTCAAGACAGAAAGAACCCACTTTTCTCTACT	615
QY	165	HisIleProTyrSerIleLeuProAlaSerValLeuLysSerGlyThrIysIleLeuAsn	184
DB	616	CATATCCAAACGGGTATTACCGGATTCGATTGTGAATCTTGTTGTGAAGATGGTGATC	675
QY	185	IleSerArgAsnArgLysSerThrPheValSerPheTrpLysPheGlyAsnLeuLeuAsn	204
DB	676	ATATGGAGACCCGCAAGATACTTTTCATCTCATGTGGACTTTC-----TTACAC	726
QY	205	ProAspLys-----LeuLeuAspLeuGluLysSerValAspIlePheAla	219
DB	727	AAGGAGAAGTCTCAAGAGGTCAATTAGCGAGTCTTGAGGACAGCTTTCATATGTTTGT	786
QY	220	SerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPheThrAsnAlaAla	239
DB	787	AAAGGTTTATCTGTATGTCCTTATCTGGATCATGTTTGGGTTTATGGAAGAGCTTAC	846
QY	240	SerThrAsn---SerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProVal	258
DB	847	CAAGAGATCCAGATAGGATTTCCTTAGGTACGAGACCATGAGGCCCAATCCTTTG	906
QY	259	GluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspGluGluLys	278
DB	907	CCTTTGTGAAGAGATTGGCTGAGTTTCATGGGTATGGATTCTACTGATCAGGAAGAGGAG	966
QY	279	GingIyIleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnGln	298
DB	967	AATGGTGTGCTGAGAAAGTGGTGAAGCTTTGTAGCTTTGAGACGTTGAAGAATCTTGA	102
QY	299	ValAsnLysAsnGlySerSerTyrAsnSerLys-----IleAspAsnLysHisPhe	315
DB	1027	GCTAACAAAGGTGATAAAGAAAGAGAGGATCGTCCTGCTGTTTATCGCAATAGCCGGTAT	108
QY	316	PheArgLysGlyGluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLysLys	335
DB	1087	TTTAGCAAGAAAGGTTGGAGTATGGCTTAATATTATTCCTCTCAGATGGCTGCTCGT	114

```

QY 336 LeuGluThrAlaGlyLysLeuGluSerGluLysHis-----LeuLeuSerAsnLys 353
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1147 ATTGAT-----GGCTATTGGAGGAGAAATCAAGATAGCTGGTCTTCAACATGAT 1200
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 354 AsnLeuThrAsnPheAsnLeuAsnLeuTyrCys-----GluLysSer 367
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1201 ACTGAATGTTGTTTATTCTGG--TCCTCTGTTATTGTCGGCGGAAACAGAGATGC 1258
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 368 AsnLeuSerLeuCysLeuLeuSerTyrSerPheValLeuAsnPheLeuLeuLys 387
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1259 TCGTTTTGCTGCTGTTGGTCTGTTATCTCATTTATGTTCTGATCATCTTATTACAA 1318
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 388 LysLysLys 390
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1319 AAGAGAAAG 1327
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 13
AY085411 1212 bp mRNA linear PLN 14-APR-2003
LOCUS Arabidopsis thaliana clone 150484 mRNA, complete sequence.
ACCESSION AY085411
VERSION AY085411.1 GI:21404121
KEYWORDS FLU CDNA.
SOURCE Arabidopsis thaliana (thale cress)
ORGANISM Arabidopsis thaliana
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsi.
1 (bases 1 to 1212)
Haas,B.J., Volfovsky,N., Town,C.D., Troughan,M., Alexandrov,N.,
Feldmann,K.A., Flavell,R.B., White,O. and Salzberg,S.L.
Full-length messenger RNA sequences greatly improve genome
annotation
Genome Biol. 3 (6), RESEARCH0029 (2002)
22088475
MEDLINE 12093376
PUBMED
REFERENCE 2 (bases 1 to 1212)
Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Full-length cDNA from Arabidopsis thaliana
Unpublished
3 (bases 1 to 1212)
Brover,V., Troughan,M., Alexandrov,N., Lu,Y.-P., Flavell,R. and
Feldmann,K.
Direct Submission
Submitted (11-MAR-2002) Ceres, Inc, 3007 Malibu Canyon Road,
Malibu, CA 90265, USA
This clone sequence is one of 5,000 Ceres full-length cDNAs made
available to TIGR and Genbank. The following quality assessment of
this set was done by comparison with known proteins: two percent of
the clones are estimated to be 5'-truncated; less than one percent
are 3'-truncated; approximately two percent represent alternative
splice variants, including unspliced introns and spliced exons; one
percent may contain premature stop codons; five percent may have
frame shifts in a coding region. A sequence is considered to be
5'-truncated if it lacks the translation initiation start (ATG). A
sequence is considered to be 3'-truncated if it lacks the
C-terminal end of the encoded protein. Please note that these cDNA
sequences are derived from the WS or LAer ecotypes and therefore
may contain polymorphisms when compared to sequences from Col-0.
Genet carried out the library production and sequencing of the
full-length clones. Ceres, Inc. carried out the clustering of the
5' sequences, selection of clones, and sequence assembly.

FEATURES             source
1..1212
/organism="Arabidopsis thaliana"
/mol_type="mRNA"
/db_xref="taxon:3702"
/clone="150484"
31..1026
/codon_start=1
/product="steroid sulfotransferase, putative"
/protein_id="AAM62638.1"

CDS

```

## ORIGIN

```

Alignment Scores:
Pred. No.: 1,35e-39 Length: 1212
Score: 572.50 Matches: 135
Percent Similarity: 56.39% Conservative: 68
Best Local Similarity: 37.50% Mismatches: 134
Query Match: 27.80% Indels: 24
DB: 8 Gaps: 9

US-09-854-122-16 (1-391) x AY085411 (1-1212)
QY 31 GluGlnGluLysGluGluAspSerLysMetTyrLysArgTyrArgGluIleValSerSer 50
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 64 GAAGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 51 LeuProSerAsn---AspTyrTyrGlyAspThrMetArgLeuTyrLysGlyPheTyrGln 69
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 124 CTTCCTTCAGACATAGATTGCTCAGGACCAAGTTGTCAAGTACCAAGATGTTGGTAC 183
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 70 MetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPheLysAlaArgGluThr 89
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 184 GACAAAGATATCTCCAGCAATCTCAATTCAACAAAACCTTCAGCCCAAGAAACG 243
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 90 AspIleLeuThrThrLeuProLysAlaGlyThrThrThrLysAlaLeuThrPhe 109
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 244 GATATAATTGTTGCTTCTTCCCAATCGGTCAGACTTGGCTTCAAGGCCACTCACATTC 303
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 110 AlalleuThrArgAspValAsnHisProSerProThrHisProLeuLeuPhePhe 129
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 304 GCATCTGGCGCAAGATCAAAACAC-----ACTTCAGACATCATCTCTGCTAACTCAT 357
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 130 AsnProHisSerCysValGlnAsnLeuGluTyr---LeuTyrMetGlyArgGluAsnThr 148
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 358 AATCCTCATGAGTAGTGCCTGACGAGTCTCCTTCAAGAGTCGTAGCTGTCAGG 411
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 149 MetProAspLeuAspMetLeu---AsnGluSerProArgLeuPheAlaGlyHisIlePro 167
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 412 AAACCGGATTTGACCAAGTTCCTCATCATCTCGAGATTGTTCTCAACCCATGTC 471
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 168 TyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIleIleAsnIleSerArg 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 472 TTTGATGCGCTTAAAGTACCCTGAAAGAGTCTCCTTCAAGAGTCGTAGCTGTCAGG 531
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 188 AsnArgLysSerThrPheValSerPheTyrLysPheGlyAsnLeuLeuAsnProAspLys 207
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 532 AACGTGAATGACGTATTGATATCACTTTGGTGTTCGAAAACCTCCATGAGTGGAGAAAC 591
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 208 LeuLeuAspLeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyPro 227
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 592 AATTTAAGTCTCGAGCTTTGTTGAGTCTTATGTCGGAGTTTAACATTATGCGTCCC 651
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 228 GluTyr---AsnPheGluAlaGluPheThrAsnAlaAlaSerThrAsnSerAsnLeuLeu 246
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 652 TTATGGAAAATGTTTAGGCTATTGGAGAGAGACCTTGGAGATCCCTAAGCATGTGCTT 711
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 247 LeuLeuSerTyrGluGluMetLysProValGluAsnValLysLysLeuAlaGlu 266
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 712 TTCCTTGAGTACGAGAGTTGAGACGGAGCTCGTGTGCAATCAAGAGATTCGAGAG 771
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 267 PheMetGlyCysGlyPheThrAspAspGluGluLysGlnGlyIleValAspGluIleVal 286
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 772 TTCTTAGATTGTCATTTCACAAAGGAGAGATAGTGGAGGTGTAGACAAAGATCTTGT 831
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 287 LysLeuCysSerPheAsnLeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyr 306
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

/db_xref="GI:21553545"
/translation="MGKDIPLRNLEKEEBEENQSBETKSLISSLPDIDCSGTYLY
KYCGWYDKDILQAILNFKNQFQDITIVAFPSKSTFWLAKALFALAKVRKHSFD
NHLLTNPHLYLYLELDLYLKSQPDITKLPSSRFLTFSTHMSFDALQVLEKSPC
KIVVCNVDNVLISLWCFNSGNNLSLEALFSLCSGVLNCPFLWENVLYVRG
SLDPKRLVFLRYBELKTEPRVOIKRLAEFLDCPFTKEEDSGGVKILGLSLRLS
GLEINKTGSLSSEGVSKFPRKGEVDKWSYMPMEMENKIDMIVBEKLGSLGLK"

```



```

Db      832 GAACCTTGGTTCTCTAAGAAACCTTAGCGGTTTGGAGATCAACAAACAGGA---AGCTTG 888
Qy      307 AsnSerLysIleAspAenLysHisPheAtrGlyGlyValArgAspTrpAlaAsn 326
      ...
Db      889 TCGAGAGAGTAAGATTCAAGAGATTTCCTCGTAAGGGAGATTGGTGATTGGAAGAT 948
Qy      327 TyrLeuThrSerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGluSerGlu 346
      ...
Db      949 TATATGACTCCTGAATGGAACAAATCGAC-----ATGATTGTTGAGGAG 996
Qy      347 LysHisLeuLeuSerAenLysAsnLeuThrAsnPheAsnLeuAsnAenTyrCysGluLys 366
      ...
Db      997 AAACCTCAAGGCTCTGTTT-GAAATTGTAGATTTC-----1031
Qy      367 SerAsnIleSerLysLeuCysLeuSerTyrSerPheValIleAsnAsnPheIleLeuLeu 386
      ...
Db      1032 -----ATATCTCTATGTATATGTGTGGGAACAGGTTTAACTCAAACTTAATATGCTG 1085

RESULT 14
AY042887
LOCUS      1221 bp mRNA linear PLN 04-JUL-2001
DEFINITION Arabidopsis thaliana putative flavonol sulfotransferase (F2P9.3)
            mRNA, complete cds.
ACCESSION      AY042887
VERSION      AY042887.1 GI:14596198
KEYWORDS      FLU CDNA.
SOURCE      Arabidopsis thaliana (thale cress)
ORGANISM      Arabidopsis thaliana
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eurosids II; Brassicales; Brassicaceae; Arabidopsids.
            1 (bases 1 to 1221)
REFERENCE      Nguyen,M., Karlin-Neumann,G., Southwick,A., Lam,B., Miranda,M.,
            Palm,C.J., Bowser,L., Jones,T., Banh,J., Carninci,P., Chen,H.,
            Cheuk,R., Chung,M.K., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Ishida,J.,
            Kawai,J., Kim,C., Lin,J., Liu,S.X., Narusaka,M., Pham,P.K.,
            Sakano,H., Sakurai,T., Satou,M., Seki,M., Shinn,P., Yamada,K.,
            Shinozaki,K., Ecker,J., Theologis,A. and Davis,R.W.
            Direct Submission
            Submitted (25-JUN-2001) DNA Sequencing and Technology Center,
            Stanford University, 855 California Avenue, Palo Alto, CA 94304,
            USA
            e-mail for correspondence: arab@sequence.stanford.edu

TITLE
JOURNAL
COMMENT

RIKEN Genomic Sciences Center (GSC) members carried out the
collection and clustering of RAFL cDNAs (RAFL CDNA : 'RIKEN
Arabidopsis Full-length cDNA'): Seki,M., Narusaka,M., Ishida,J.,
Satou,M., Kamiya,A., Sakurai,T., Carninci,P., Kawai,J.,
Hayashizaki,Y. and Shinozaki,K.

The Salk, Stanford, PGEN (SSP) Consortium members carried out the
sequencing and annotation of the RAFL cDNAs: Nguyen,M.,
Southwick,A., Karlin-Neumann,G., Lam,B., Miranda,M., Palm,C.J.,
Bowser,L., Jones,T., Banh,J., Chen,H., Cheuk,R., Chung,M.K.,
Kim,C., Lin,J., Liu,S.X., Pham,P.K., Sakano,H., Shinn,P.,
Yamada,K., Ecker,J., Theologis,A. and Davis,R.W.

Nguyen,M., (SSP/Stanford) and Seki,M. (RIKEN GSC) contributed
equally to this work. Shinozaki,K. (RIKEN GSC) and Davis,R.W.
(SSP/Stanford) contributed equally to this work as PIs.

FEATURES
source      1..1221
            /organism="Arabidopsis thaliana"
            /mol_type="mRNA"
            /db_xref="taxon:3702"
            /chromosome="1"
            /clone="RAFL05-13-F01"
            /ecotype="Columbia"
            /note="This clone is in pBluescript"
            1..1221
            /gene="F2P9.3"
            76..1092
            /gene="F2P9.3"
            /gene="F2P9.3"

```

```

/codon_start=1
/product="putative flavonol sulfotransferase"
/protein_id="RAK6827_1"
/db_xref="GI:14596198"
/translation="MESKTONGSEVVELTEFEKTKYQDFIATLPKSKGWPRDEL
TQGHMWECELEGLFHAKDHFARPFDLVCSPKGTITWTKALTYLVNRSYDD
AANPLKNPHEFVYVEIDFAFYTDVQDRKNPLFTSHIPLGLPDSLVNNGCQD
VYIWDPKDTIFISMTWFLHKSQEGGLASLEDFMFCGLSVLYGLDHLVGYWKA
YOENDRILFYETMRANPLFPVKRLAEFMGYGTDEEENGVAEYVKLCLSPETLK
NLEANKGDKEREDRPVAVYANGAYFRKGVKGVANLYLTPMAARIDGLVEKFKDTGLL
QHDN"

ORIGIN
Alignment Scores:
Pred. No.:      2,238-39      Length:      1221
Score:          570.00      Matches:      140
Percent Similarity: 53.53%      Conservative: 69
Best Local Similarity: 35.90%      Mismatches: 144
Query Match:      27.68%      Indels:      38
DB:              8      Gaps:      12

US-09-854-122-16 (1-391) x AY042887 (1-1221)
Qy      14 LeuIleMetAlaGlyIleLeuAlaLeuGluLysCysPheGlySerLysAsnGluIn--- 32
      ...
Db      40 ATCTCTTTGACTCCCTTCACTCGATCTCATCCGAATGGAATCAAAAGACAAACCAAAAC 99
Qy      33 -----GluLysGluGluAspSerLysMetTyrLysAsnGlyTrpArgGluLe 47
      ...
Db      100 GGATCGGAAGTCGTGCACTACAGAGTTCGAGAAACCCAGAGAGATATCAAGATTTC 159
Qy      48 ValSerSerLeuProSerAsnAspTyrTrp-----GlyAspThrMetArgLeuTyrLys 65
      ...
Db      160 ATCGGTACACTTCCAAAGAGCAAAAGCGTGGAGACAGATGAGATCTTAAACCAATACGGT 219
Qy      66 GlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPhelys 85
      ...
Db      220 GGACACTGGTGGCAAGAATGCTCTCGAAGTCTTTTTCAGCTAAAGACCAATTCGAA 279
Qy      86 AlaArgGluThrAspIleLeuThrThrLeuProLysAlaGlyThrThrThrThrLys 105
      ...
Db      280 GACGACCACTAGATTTCTCTGCTGTAGCTACCCAAACCCGCTACCACTGGCTCAA 339
Qy      106 AlaLeuThrPheAlaIleLeuThrArgAspValAsnHisProSerSerProThrHisPro 125
      ...
Db      340 GCACCTAACCTTACGCAATCGTCAATCGT-----TCTCGTACGACGACGCGCAACCCA 393
Qy      126 LeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArg 145
      ...
Db      394 CTCCTCAACAGCAACCCCTCACAGAGTTTGTCCCTTACGTTGAGATCGACTTCGGC----- 447
Qy      146 GluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArg---LeuPheAlaGly 164
      ...
Db      448 -----TTTATACCCACCGTTGATGTTCTTCAAGACAGAGAAAGAACCCACITTTCTACT 501
Qy      165 HisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIleAsn 184
      ...
Db      502 CATATCCCAACAGGTTTATACCCGATTCGATTCGAACTCTGTTGTAAGATGGTGATC 561
Qy      185 IleSerArgAsnArgLysSerThrPheValSerPheTrpLysPheGlyAsnLeuIleAsn 204
      ...
Db      562 ATATCGAGAGACCCGCAAGATACITTCATCTCCATGTCGACTTTC-----TTACAC 612
Qy      205 ProAspLys-----LeuLeuAspLeuLysSerValAspIlePheAla 219
      ...
Db      613 AAGGAGAGAGTCTCAAGAGGTCATATACCGAGTCTTGAGGACAGCTTTGATATTTTGT 672
Qy      220 SerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPheThrAsnAlaAla 239
      ...
Db      673 AAAGGTTTATCTGCTGATGTCCTATCTGATCATGTTTGGGTATTTCGAAGCTTAC 732
Qy      240 SerThrAsn---SerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProVal 258
      ...

```

```

Db      733 CAAGAGATCCAGATAGGATTTTGTCTTCTAGGTACGACCATGAGGCCCAATCCTTTG 792
Qy      259 GluAsnValLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluLys 278
Db      793 CCTTTGTGAAGAGATTGCTGAGTTTCATGGGTTATGGAATTCATGATGAGGAGGAG 852
Qy      279 GlnGlyIleValAspGluLysLeuValLysLeuCysSerPheAspAsnLeuLysAsnGln 298
Db      853 ATGGTGTGCTGAGAAAGTGTGAAGCTTTGAGCTTTGAGACCTTGAAGATCTTGA 912
Qy      299 ValAsnLysAsnGlySerSerTyrrAsnSerLys-----IleAspAsnLysHisPhe 315
Db      913 GCTAACAAAGGTCATAAAGAAAGAGAGGATCGTCTGCTGTTTATGCGAATAGCCGAT 972
Qy      316 PheArgLysGlyGluValArgAspTrpAlaAsnTyrrLeuThrSerGluMetIleLysLys 335
Db      973 TTTAGAAAGGAAGTTGGAGTTGGGCTAATATTGACTTCCTGAGATGGCTCGT 1032
Qy      336 LeuGluThrAlaGlyLysIleAsnGluSerGluLysHis-----LeuLeuSerAsnLys 353
Db      1033 ATTGAT-----GGCTTAGTGAGGAGAAATTCAMAGATACCTGGCTTGCTTCAACATGAT 1086
Qy      354 AsnLeuThrAsnPheAsnLeuAsnAsnTyrrCys-----GluLysSer 367
Db      1087 AACTGAATGTGTTTATCTCGG--TCCTTCTGTTATTGGTGGCGGAAACAGAGATGC 1144
Qy      368 AsnIleSerLeuCysLeuLeuSerTyrrSer 377
Db      1145 TCTGTTTGTGTTTGTGTCAGTATCT 1174

RESULT 15
AC146342/c
LOCUS   Medicago truncatula clone mth2-10c20, WORKING DRAFT SEQUENCE, 2
DEFINITION
unordered pieces.
ACCESSION AC146342
VERSION    4
KEYWORDS   HTG; HTGS PHASE1; HTGS DRAFT.
SOURCE     Medicago truncatula (barrel medic)
ORGANISM   Medicago truncatula
            Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
            Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
            rosids; eutrosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
            Medicago.
REFERENCE  1 (bases 1 to 113531)
            Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
            Medicago truncatula BAC Clone mth2-10c20
            Unpublished
REFERENCE  2 (bases 1 to 113531)
            Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
            Direct Submission
            Submitted (12-AUG-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
REFERENCE  3 (bases 1 to 113531)
            Lin,S., Dixon,R., May,G., Sumner,L., Gonzales,B., Cook,D., Kim,D.
            and Roe,B.A.
            Direct Submission
            Submitted (21-NOV-2003) Department Of Chemistry And Biochemistry,
            The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,
            OK 73019, USA
COMMENT   On Nov 21, 2003 this sequence version replaced gi:38371842.
            ----- Genome Center.
            Center: Department Of Chemistry And Biochemistry
            The University Of Oklahoma
            Center code:UOKNOR

```

\* NOTE: This is a 'working draft' sequence. It currently consists of 2 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as

\* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

1 27561: contig of 27561 bp in length  
 27562 27661: gap of unknown length  
 27662 113531: contig of 8570 bp in length.

FEATURES  
 source  
 1..113531  
 /organism="Medicago truncatula"  
 /mol\_type="genomic DNA"  
 /db\_xref="taxon:3880"  
 /clone="mth2-10c20"  
 /clone\_lib="Medicago truncatula BAC library H2"

## ORIGIN

Alignment Scores:  
 Pred. No.: 6.87e-37 Length: 113531  
 Score: 567.50 Matches: 125  
 Percent Similarity: 57.27% Conservative: 64  
 Best Local Similarity: 37.88% Mismatches: 122  
 Query Match: 27.56% Indels: 19  
 DB: 2 Gaps: 10

US-09-854-122-16 (1-391) x AC146342 (1-113531)

```

Qy      37 AspSerLysMetTyrrLysArgTyrrArgGluIleValSerSerLeuPro----SerAsnAsp 55
Db      109250 GAAACAAGGTAATCCATGATTATAACCTAATCTTGTCACTTCCTAGAGAAATGGG 109191
Qy      56 TyrTrpGlyAspThrMetArgLeuTyrrLysGlyPheTrpGlnMetGlyTyrrLeuValPro 75
Db      109190 TGTGAGACTGTAATATATGATTTTTCATGGATTTTGGTGTCATCAATCTTATTCCAA 109131
Qy      76 GlyIleMetAlaPheGluAspAsnPheLysAlaArgGluThrAspIleLeuLeuThrThr 95
Db      109130 TCTGTCAACACTTTTCAAAATAAATTTTCATGCCAAGACAATGATATTTGTTGCATCT 109071
Qy      96 LeuProLysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAsp 115
Db      109070 ATGCCAAATCTGGCACCACTTGCTTAAAGGCTTCATATGCTATTGTTAATCGC--- 109014
Qy      116 ValAsnHisProSerSer-----ProThrHisProLeuLeuPhePheAsnProHis 132
Db      109013 ---CAACATTTTACCTCTCTTAGAGAAATAAACCATCCATCTATTATTATATCCACAT 108957
Qy      133 SerCysValGlnAsnLeuGluTyrr---LeuTyrrMetGlyArgGluAsnThrMetProAsp 151
Db      108956 GAACCTAGTGCCTCTGTTTGAAGGTAAACCTTTATGGTGGTAAAGACATCTTTTGGCCTCAA 108897
Qy      152 LeuAspMetLeuAsnGlu---SerProArgLeuPheAlaGlyHisIleProTyrrSerLeu 170
Db      108896 ATTGATGTAATCAATATGATTGAACCAAGGCTTTTGGCACTCACATTCCTCTTCG 108837
Qy      171 LeuProLysSerValLeuLysSerGlyThrLysIleIleAsnIleSerArgAsnArgLys 190
Db      108836 TTGCCCAAGTCGGTTAAAGAGTCCAGTTGAAATCAATTTATATATGTAAGAACCCCAATTT 108777
Qy      191 SerThrPheValSerPheTrpLysPheGlyAsnLeuIleAsnProAspLysLeu----- 208
Db      108776 CATACATTTGTATCATATTGGAATTATCAACAAGCTCAGATCGAAGAATCTTTAACC 108717
Qy      209 ---LeuAspLeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyPro 227
Db      108716 GAAATTAACCTTAGAGGAGTCTTTTGAAGGTAATTTGTAAGGGAATATGTTATTATTGGTCCG 108657
Qy      228 GluTrp---AsnPheGlnAlaGluPheThrAsnAlaAlaSerThrAsnSerAsnLeuLeu 246
Db      108656 TTTTGGGACAATATGTTGGGTACTTCTGAAGAGAGTATAGAAGACACACATAGGTTTAA 108597
Qy      247 LeuLeuSerTyrrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGlu 266
Db      108596 TTCITGAAATATGAAGACCTTAAAGAGATGTTAATTTTACATCAAAAGAAATTCAGAG 108537

```



```
Qy 267 PheMetGlyCysGlyPheThrAspGluGluLysGlnGlyIleValAspGluIleVal 286
Db 108536 TTTGTGGATTCTTTCATCTCAGAGAGAGAAATAATGGAGTGGTTGAAATATAATC 108477
Qy 287 LysLeuCysSerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyr 306
Db 108476 AAGTTATGCAGCTTTCAGAGTATGAGGAATCAAGGAAATCAATCT--GAAACTGTA 108420
Qy 307 AsnSerLysIleAspAsnLysHisPheArgLysGlyGluValArgAspTTrpAlaAsn 326
Db 108419 TCACAAATCTATGAGAAAGTTTCTTTCGAAAGGAGAAATAGAGATTGGGTAAT 108360
Qy 327 TyrLeuThrSerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGluSerGlu 346
Db 108359 TACTTTCTCCTTCAATGATAGAAAGATTA-----TCCAAAGTCATAGAGAAATA 108309
Qy 347 LysHisLeuLeuSerAsnLysAsnLeuThr 356
Db 108308 -----TTAGTGGATCAACCTATCA 108288
```

Search completed: March 20, 2004, 05:31:28  
Job time : 5431 secs

GenCore version 5.1.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 20, 2004, 01:48:09 ; Search time 507 Seconds

(without alignments)  
3276.224 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059  
Sequence: 1 TRGITGIAVACPLIMAGIL.....CLLSYFVNNFILLKKKK 391

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 3373863 seqs, 212409041 residues

Total number of hits satisfying chosen parameters: 6747726

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODE=frame\_plus\_p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09854122/runat\_18032004\_100000\_3247/app\_query.fasta\_1.583  
-DB=N Geneseq 29Jan04 -QFMT=fastap -SUFFIX=ring -MINMATCH=0.1 -LOOPEL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DICALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854122.cgn\_1\_1\_470 @runat\_18032004\_100000\_3247 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 29Jan04:\*  
1: Geneseqn1980s:\*  
2: Geneseqn1990s:\*  
3: Geneseqn2000s:\*  
4: Geneseqn2001as:\*  
5: Geneseqn2001bs:\*  
6: Geneseqn2002as:\*  
7: Geneseqn2003as:\*  
8: Geneseqn2003bs:\*  
9: Geneseqn2003cs:\*  
10: Geneseqn2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Query Match	Score	Length	ID	Description
1	1996	96.9	1192	6 AAS20863	Aas20863 Z. marina
2	573	27.8	1347	3 Aac37555	Aac37555 Arabidops
3	572.5	27.8	1212	3 Aac39344	Aac39344 Arabidops
4	567	27.5	1041	7 Abz68953	Abz68953 Nucleotid
5	566	27.5	1281	3 Aac37716	Aac37716 Arabidops
6	562.5	27.3	1017	7 Abz68951	Abz68951 Nucleotid
7	558.5	27.1	1053	7 Abz68952	Abz68952 Nucleotid
8	558	27.1	1400	7 Aad54410	Aad54410 Lolium pe

9	557.5	27.1	1205	3 AAC37521	Aac37521 Arabidops
10	555.5	27.0	1077	4 Aaf29177	Aaf29177 HydroxyJa
11	555.5	27.0	1270	3 Aac48839	Aac48839 Arabidops
12	555.5	27.0	1273	3 Aac47848	Aac47848 Arabidops
13	548.5	26.6	1154	3 AAC37672	Aac37672 Arabidops
14	545	26.5	936	6 ABZ13222	Abz13222 Arabidops
15	537.5	26.1	980	7 ADA68055	Ada68055 Arabidops
16	537.5	26.1	981	6 ABZ14244	Abz14244 Arabidops
17	537.5	26.1	981	7 ABZ42036	Abz42036 Arabidops
18	537.5	26.1	1043	3 AAC42382	Aac42382 Arabidops
19	528.5	25.7	999	7 ADA69659	Ada69659 Rice gene
20	527	25.6	271990	9 ADD25213	Add25213 Fertility
21	521.5	25.3	1017	7 ADA69350	Ada69350 Rice gene
22	515.5	25.0	1160	3 AAC32748	Aac32748 Arabidops
23	510.5	24.8	1041	4 AAF29178	Aaf29178 HydroxyJa
24	483	23.5	2067	9 ADD25206	Add25206 Fertility
25	479	23.3	1032	7 ADA71112	Ada71112 Rice gene
26	446.5	21.7	1040	7 ADA71367	Ada71367 Rice gene
27	443.5	21.5	1460	7 ADA70067	Ada70067 Rice gene
28	409	19.9	776	7 Aad54404	Aad54404 Lolium pe
29	396	19.2	776	7 Aad54402	Aad54402 Lolium pe
30	372	18.1	777	7 Aad54405	Aad54405 Lolium pe
31	338.5	16.4	1000	8 ADB23209	Adb23209 Environme
32	330	16.0	2000	7 ADA71733	Ada71733 Rice gene
33	321	15.6	515	9 ADE81349	Ade81349 Arabidops
34	317.5	15.4	668	7 Aad54403	Aad54403 Lolium pe
35	273	13.3	1063	2 AAV11154	Aav11154 Human EST
36	273	13.3	1063	2 AAV12466	Aav12466 Human oes
37	270.5	13.1	566	7 ABX56987	Abx56987 Arabidops
38	268	13.0	1004	4 AAD19778	Aad19778 Human ST
39	268	13.0	1240	4 AAD09944	Aad09944 Human dru
40	264	12.8	942	4 AAD19776	Aad19776 Human ST
41	262	12.7	1600	6 ABN83170	Abn83170 Human sul
42	260	12.6	915	6 ABN83168	Abn83168 Human sul
43	260	12.6	922	4 AAD19777	Aad19777 Human ST
44	259.5	12.6	849	6 ABN83171	Abn83171 Human sul
45	257	12.5	1363	6 ABK63484	Abk63484 Rat seque

#### ALIGNMENTS

RESULT 1

AAS20863  
ID AAS20863 standard; cDNA; 1192 BP.

XX XX AAS20863;

XX XX 09-APR-2002 (first entry)

DE DE Z. marina cDNA clone encoding protein containing sulfotransferase.

XX XX Plant; transgenic; marine eelgrass; zosteric acid biosynthesis;  
XX KW saline-resistance; anoxia-resistance; anti-fouling genetic trait;  
KW marine vascular plant; sulphated phenolic compound; Zostera marina;  
KW sulfotransferase; ST; enzyme; gene; ss.

XX XX Zostera marina.

XX XX Key Location/Qualifiers

FT CDS 1..1192

FT /tag= a

FT /partial

FT /product= "Protein containing sulfotransferase"

FT /note= "This sequence lacks both start and stop codons"

FT /transl\_except= (pos:34..36, aa:Xaa)

FT /transl\_except= (pos:1042..1044, aa:Xaa)

FT /transl\_except= (pos:1060..1062, aa:Xaa)

FT /transl\_except= (pos:1078..1080, aa:Xaa)

FT /transl\_except= (pos:1087..1089, aa:Xaa)

FT /transl\_except= (pos:1120..1122, aa:Xaa)

FT /note= "Xaa= Stop codon"

XX PN W0200185971-A2.

XX 15-NOV-2001.  
XX 10-MAY-2001; 2001WO-US015412.  
XX 10-MAY-2000; 2000US-0202529P.  
XX (PHYC-) PHYCOGEN INC.  
XX Alberte RS, Smith RD;  
XX WPI; 2002-121947/16.  
XX P-PSDB; AAU12056.  
XX New transgenic plants comprising a zosteric acid biosynthetic gene, a  
XX salin resistance gene or a hypoxia resistance gene derived from Zostera  
XX marina, useful for producing plants with antifouling traits.  
XX Example; Fig 4; 117pp; English.  
XX The present invention relates to a new transgenic plant comprising a  
XX heterologous gene derived from the marine eelgrass Zostera marina or at  
XX least one heterologous nucleotide sequence encoding a zosteric acid  
XX biosynthetic function, a salin-resistance function, or a anoxia-  
XX resistance function. The invention describes the method of producing a  
XX transgenic plant possessing an anti-fouling genetic trait by providing a  
XX cDNA population derived from a marine vascular plant, isolating from the  
XX cDNA population a nucleic acid species which hybridises to a nucleic acid  
XX that encodes a sulfotransferase (ST), an alcohol dehydrogenase (ADH), and  
XX phenylalanine ammonia lyase (PAL) or a cinnamate-4-hydroxylase (CH), and  
XX transforming a target host plant with the isolated nucleic acid. The  
XX plant is useful in the genetic engineering of plant species having  
XX desirable genetic traits such as antifouling traits, salt and anoxia  
XX resistance, and pathogen defence strategy. The expression of such  
XX biosynthetic enzymes are sufficient to support the production of zosteric  
XX acid and other sulphated phenolic compounds in a target plant. The  
XX present sequence represents a Z. marina cDNA clone which encodes a  
XX protein containing sulfotransferase  
XX  
SQ Sequence 1192 BP; 391 A; 206 C; 266 G; 329 T; 0 U; 0 Other;  
Alignment Scores:  
Pred. No.: 2,32e-185 Length: 1192  
Score: 1996.00 Matches: 391  
Percent Similarity: 98.49% Conservative: 0  
Best Local Similarity: 98.49% Mismatches: 0  
Query Match: 96.94% Indels: 6  
DB: 6 Gaps: 6  
US-09-854-122-16 (1-391) x AAS20863 (1-1192)  
QY 1 ThrArgGlyIleThrGlyIleAlaValAlaCys---LeuProLeuIleMetAlaGlyIle 19  
DB 1 ACGCGGGGAATAACTGGGAATCGCTGTTGCTGCTAGCTACCACCTGATTAATGGCTGGAAAT 60  
QY 20 LeuAlaLeuGluLysCysPheGlySerLysAsnGluGlnGluLysGluGluAspSerLys 39  
DB 61 TTAGCTTTGGGAATATGTTCCGATCCGAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
QY 40 MetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTyrGlyAsp 59  
DB 121 ATGTACAAGAGATATAGAGAGATGTTTCTTCACCTCCCTCCCTCGAATGATATATGGGGGAT 180  
QY 60 ThrMetArgLeuTyrLysGlyPheTyrGlnMetGlyTyrLeuValProGlyIleMetAla 79  
DB 181 ACCATGAGGTTGTACAGGGATTTGGCAATGGGATATCTGTACCTGGTATCATGGCT 240  
QY 80 PheGluAspAsnPhelLysAlaArgGluThrAspIleIleLeuThrThrLeuProLysAla 99  
DB 241 TTCGAAGATAATTTCAAGGCTCGAGAGACGACATATCTTACGACTCTTCCAAAGGCT 300  
QY 100 GlyThrThrTyrThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisPro 119

DB 301 GGAACGACATGGACGAAGGCACTGACGTTTGCCATCTAAACACGAGATGTTAACCCACA 360  
QY 120 SerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGlu 139  
DB 361 TCATCACCGACACATCCACTTTTCTTCAACCTCATTCGTGTGTTCAAAATTTGGAG 420  
QY 140 TyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerPro 159  
DB 421 TATTTGTACATGGGTAGAGAAATACGATGCCAGCCTCGATATGTTGAATGAAATCGCG 480  
QY 160 ArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGly 179  
DB 481 AGGTTGTTTGGCGACACATCCCATCTTTTGTGGCGGCTCTGTTTGAATTCGGGA 540  
QY 180 ThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPheTyrLysPhe 199  
DB 541 ACAAAATCATCAATATAAGCCGACACCGTAAGAGTACATTTGTCTTTTGGAAATTT 600  
QY 200 GlyAsnLeuIleAsnProAspLysLeuAspLeuLysSerValAspIlePheAla 219  
DB 601 GGCATCTGATTAAACCCGACAGTTATTGGACCTCGAAAGAGCGTTGATATCTTCGA 660  
QY 220 SerGlyIleSerPheCysGlyProGluTyrAsnPheGlnAlaGluPheThrAsnAlaAla 239  
DB 661 TCGGGAATCTCTCTTTTGGACCGAATGGATTTCCAAAGCGAGTTCCACCAATCGCGC 720  
QY 240 SerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGlu 259  
DB 721 TCTACTAATTCAAACTTCTGCTATTGTTAGTTTACGAAGAAATGTTAGAGAAGCCAGTTGAA 780  
QY 260 AsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluLysGln 279  
DB 781 AATGTGAGAGAGCTAGCTAGTTCTATGGGATGTTGGTTTCAGACGATGAGGAGAAACAA 840  
QY 280 GlyIleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnVal 299  
DB 841 GCGAATGTTGATGAGATAGTTAAACTTTTGTAGCTTCGACAAATCTGAAGAATCAACAGGT 900  
QY 300 AsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHisPheArgLysGly 319  
DB 901 AACAAACCGGATCAAGCTCAATTCGAATTCGACACACAGCATTTCTTCAGGAAGGT 960  
QY 320 GluValArgAspTyrAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGluThrAla 339  
DB 961 GAGGTGAGAGATGGGCAAACTATCTAAACGTCGGAATGATTAGAAAACGAGACGGCC 1020  
QY 340 GlyLysIleAsnGluSerGlu---LysHisLeuLeuSer---AsnLysAsnLeuThr--- 356  
DB 1021 GGAATAATAATGAATCAGAGTAAAGCATTTTATTCTGTGAATAGAAATCTTACATGA 1080  
QY 357 AsnPhe---AsnLeuAsnAsnTyrCysGluLysSerAsn---IleSerLeuCysLeuLeu 374  
DB 1081 AACTTCTGAAATCTTAATAATTAATCTGTGGAATTCGAATAATATCTTTGTTTATTA 1140  
QY 375 SerTyrSerPheValIleAsnAsnPhelIleLeuLeuLysLysLysLys 391  
DB 1141 TCGTATTCATTCGTAATAATAATTTTCATTTGTTTAAAAAATAAAAAA 1191  
RESULT 2  
AAC37555  
ID AAC37555 standard; DNA; 1347 BP.  
XX AAC37555;  
XX AAC37555;  
DT 17-OCT-2000 (first entry)  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 17816.  
XX Hybridisation assay; genetic mapping; gene expression control;  
XX protein identification; signal transduction pathway; metabolic pathway;  
XX promoter; termination sequence; ss.  
OS Arabidopsis thaliana.

XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX PP 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131149P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134376P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135829P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139859P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145132P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.

[illegible]

KW Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway; metabolic pathway;  
XX promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
XX EP1033405-A2.  
XX 06-SEP-2000.  
XX  
XX 25-FEB-2000; 2000EP-00301439.  
XX  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0126264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127462P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 18-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 28-APR-1999; 99US-0130891P.  
XX 30-APR-1999; 99US-0131449P.  
XX 01-MAY-1999; 99US-0132048P.  
XX 04-MAY-1999; 99US-0132407P.  
XX 05-MAY-1999; 99US-0132484P.  
XX 06-MAY-1999; 99US-0132485P.  
XX 07-MAY-1999; 99US-0132486P.  
XX 11-MAY-1999; 99US-0132863P.  
XX 14-MAY-1999; 99US-0134256P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134221P.  
XX 14-MAY-1999; 99US-0134370P.  
XX 18-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135353P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0136021P.  
XX 27-MAY-1999; 99US-0136392P.  
XX 28-MAY-1999; 99US-0136782P.  
XX 01-JUN-1999; 99US-0137222P.  
XX 03-JUN-1999; 99US-0137528P.  
XX 07-JUN-1999; 99US-0137724P.  
XX 08-JUN-1999; 99US-0138094P.  
XX 10-JUN-1999; 99US-0138540P.  
XX 10-JUN-1999; 99US-0138847P.  
XX 14-JUN-1999; 99US-0139119P.  
XX 16-JUN-1999; 99US-0139452P.  
XX 16-JUN-1999; 99US-0139453P.  
XX 17-JUN-1999; 99US-0139492P.  
XX 18-JUN-1999; 99US-0139454P.  
XX 18-JUN-1999; 99US-0139455P.  
XX 18-JUN-1999; 99US-0139456P.  
XX 18-JUN-1999; 99US-0139457P.  
XX 18-JUN-1999; 99US-0139458P.  
XX 18-JUN-1999; 99US-0139459P.  
XX 18-JUN-1999; 99US-0139460P.  
XX 18-JUN-1999; 99US-0139461P.  
XX 18-JUN-1999; 99US-0139462P.  
XX 18-JUN-1999; 99US-0139463P.  
XX 18-JUN-1999; 99US-0139750P.  
XX 18-JUN-1999; 99US-0139763P.  
XX 21-JUN-1999; 99US-0139817P.  
XX 22-JUN-1999; 99US-0139859P.  
XX 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 22-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145182P.  
PR 22-JUL-1999; 99US-0145183P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.  
PR 17-AUG-1999; 99US-0149195P.  
PR 18-AUG-1999; 99US-0149486P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.







PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134376P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137232P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139432P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 28-JUN-1999; 99US-0140823P.  
PR 29-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143572P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 05-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 06-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148319P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0148388P.  
PR 17-AUG-1999; 99US-0148175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150684P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152263P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0158458P.  
PR 29-SEP-1999; 99US-0158596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0159369P.  
PR 13-OCT-1999; 99US-0159293P.



PA (UYCO-) UNIV CONCORDIA.  
 XX Varin L, Spertini D;  
 XX WPI; 2003-248082/24.  
 DR P-PSDB; ABP72962.  
 XX Novel isolated or purified polypeptide having biological activity of  
 PT desulfoglucosinolate sulfotransferases, useful as enzymes involved in  
 PT glucosinolate biosynthesis in transgenic plants.  
 XX Claim 9; Page 54-55; 68pp; English.  
 XX The present sequence encodes a plant desulfoglucosinolate  
 CC sulfotransferase, designated AtST5a. The specification also describe  
 CC AtST5b and AtST5c. AtST5a and AtST5b are constitutively expressed in all  
 CC parts of the plants, while AtST5c does not seem to be expressed in the  
 CC early stage of development. These enzymes are involved in glucosinolate  
 CC synthesis in plants. The desulfoglucosinolate sulfotransferase  
 CC polynucleotides are useful for modulating glucosinolate biosynthesis in  
 CC transgenic plants  
 XX Sequence 1017 BP; 296 A; 205 C; 240 G; 276 T; 0 U; 0 Other;  
 SQ

Alignment Scores:  
 Pred. No.: 2,68e-45 Length: 1017  
 Score: 562.50 Matches: 129  
 Percent Similarity: 55.95% Conservative: 59  
 Best Local Similarity: 38.39% Mismatches: 123  
 Query Match: 27.32% Indels: 25  
 DB: Gaps: 10

US-09-854-122-16 (1-391) x ABZ68951 (1-1017)

QY 33 GluLysGluLysSerLysMetLysArgLysGluLysValSerSerLeuPro 52  
 DB 40 GAACCTCAGAGCTTCGAGAAACCCAGAGAGAGTATCAAGATTTCATCGTACACTTCCA 99  
 QY 53 SerAsnAspTyrTrp-----GlyAspThrMetArgLeuTyrLysGlyPheTrpGlnMet 70  
 DB 100 AAGAGCAAGGCTGGAGACCATGAGATCTTAACCAATACGGTGGACACTGGTGGCAA 159  
 QY 71 GlyTyrLeuValProGlyLysMetAlaPheGluAspAsnLysPheLysAlaArgGluThrAsp 90  
 DB 160 GAATGTCTCCTCGAAGGTCTTTTTCACGCTAAAGACCATCTTCGAAGCAGCAGCAACTGAT 219  
 QY 91 IleIleLeuThrThrLeuProLysAlaGlyThrThrLysLysAlaLeuThrPheAla 110  
 DB 220 TTCCTCGTGTGAGTACCAACCAACCGGTACAACTTGGCTCAAGCAGCACTTACCGCA 279  
 QY 111 IleLeuThrArgAspValAsnHisProSerProThrHisProLeuLeuPhePheAsn 130  
 DB 280 ATCGTCAATCGT-----TCTCGTTACGACGACGCGCAACCCACTCTCTCAAGCAAC 333  
 QY 131 ProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetPro 150  
 DB 334 CPTCAGAGTTTCTCCCTACGTTGAGATCGACTTCGG-----TTTTTACCCC 381  
 QY 151 AspLeuAspMetLeuAsnGluSerProArg---LeuPheAlaGlyHisIleProTyrSer 169  
 DB 382 ACCGTTGATGTTCTTCAAGACAGAAAGAACCCACTTTTCTACTACTATCCCAACCGG 441  
 QY 170 LeuLeuProAlaSerValLeuLysSerGlyThrLysLysIleAsnLysSerArgAsnArg 189  
 DB 442 TTATTACCCGATTCGATTGGAATCTGGTTGTAAGATGGTGTACATATGAGAGACCGG 501  
 QY 190 LysSerThrPheValSerPheTrpLysPheGlyAsnLeuLeuAsnProAspLys----- 207  
 DB 502 AAGATACITTCATCTCCATGTCGACTTC-----TTACACAGAGAGAGTCTCAA 552  
 QY 208 -----LeuLeuAspLeuGluLysSerValAsnLysPhePheAlaSerGlyIleSerPhe.224  
 DB 553 GAAGGTCAATTAGCGAGTCTTGAGGACAGCTTTGATATGTTTGTAAAGGTTTATCTGNG 612

QY 225 CysGlyProGluTyrAsnPheGlnAlaGluPheThrAsnAlaIleSerThrAsn---Ser 243  
 DB 613 TATGTCCTTATCTGGATCATGTTTGGGTATTGGAAAGCTTACCAAGAGATCCAGAT 672  
 QY 244 AsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnValLysLys 263  
 DB 673 AGGATTTTGTCTTAGGTACGAGACCATGAGGGCCAATCTTTGCCCTTTGTGAAGAGA 732  
 QY 264 LeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGlnGlyIleValAsp 283  
 DB 733 TTGGCTGATTCATGGGTTATGGATTCTGATGAGGAAGAGGAGAAATGGTGTCTCTGAG 792  
 QY 284 GluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnValAsnLysAsnGly 303  
 DB 793 AAGTGGTGAAGCTTTGAGCTTTGAGAGCGTTGAGAAATCTTGAAGCTAACAAAGGTGAT 852  
 QY 304 SerSerTyrAsnSerLys-----IleAspAsnLysHisPhePheArgLysGlyGlu 320  
 DB 853 AAAGAAGAGAGAGATGCTCTCTGCTCTTTATGCAATAGCGCGTATTTTGAAGAAAGAAAG 912  
 QY 321 ValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGluThrAlaGly 340  
 DB 913 GTTGGAGATTGGCTAAATTATTTGACTCTGAGATGGCTGCTCGATTGAT-----GGC 966  
 QY 341 LysIleAsnGluSerGluLysHis-----LeuLeuSerAsnLysAsn 354  
 DB 967 TTAGTGGAGAGAAATTCAGAGATACCTGGCTTGTCTTCAACATGATAAC 1014  
 XX ABZ68952  
 XX ABZ68952 standard; DNA; 1053 BP.  
 XX AC ABZ68952;  
 XX 28-MAY-2003 (first entry)  
 DT Nucleotide sequence of desulfoglucosinolate sulfotransferase AtST5b.  
 XX Plant; desulfoglucosinolate sulfotransferase; AtST5a; AtST5b; AtST5c;  
 XX glucosinolate; transgenic plant; gene; ss.  
 XX Arabidopsis thaliana.  
 XX Key Location/Qualifiers  
 FH CDS 1..1053  
 FT /tag= a  
 FT /product= "desulfoglucosinolate sulfotransferase AtST5b"  
 XX W02003010318-A2.  
 XX 06-FEB-2003.  
 XX 24-JUL-2002; 2002WO-CA001144.  
 XX 24-JUL-2001; 2001US-0307141P.  
 XX (UYCO-) UNIV CONCORDIA.  
 XX Varin L, Spertini D;  
 XX WPI; 2003-248082/24.  
 DR P-PSDB; ABP72963.  
 XX Novel isolated or purified polypeptide having biological activity of  
 PT desulfoglucosinolate sulfotransferases, useful as enzymes involved in  
 PT glucosinolate biosynthesis in transgenic plants.  
 XX Claim 9; Page 57-59; 68pp; English.  
 XX The present sequence encodes a plant desulfoglucosinolate  
 CC sulfotransferase, designated AtST5b. The specification also describes  
 CC AtST5a and AtST5c. AtST5a and AtST5b are constitutively expressed in all

CC parts of the plants, while AtST5c does not seem to be expressed in the  
 CC early stage of development. These enzymes are involved in glucosinolate  
 CC synthesis in plants. The desulfoglucosinolate sulfoxidase  
 CC polynucleotides are useful for modulating glucosinolate biosynthesis in  
 CC transgenic plants

XX SQ Sequence 1053 BP; 292 A; 235 C; 259 G; 267 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 6.9e-45 Length: 1053  
 Score: 558.50 Matches: 129  
 Percent Similarity: 56.80% Conservative: 59  
 Best Local Similarity: 38.97% Mismatches: 120  
 Query Match: 27.12% Indels: 23  
 DB: 7 Gaps: 5

US-09-854-122-16 (1-391) x ABZ68952 (1-1053)

QY 29 LysAsnGluGlnGluLysGluLysGluLysMetTyrLysArgTyrArgGluLeuVal 48  
 DB 64 GAGACGACAGACAGATCAACAGAGTTCGAGAAATCAAAACGGTATCAAGACCTCATC 123  
 QY 49 SerSerLeuProSerAsnAspTyrTyrGlyAspThrMetArgLeu-----TyrLysGly 66  
 DB 124 TCCACGTTTCTCCACGAGAAGCTGAGACCGAAGAGCCCTGATCGAGTATGTGTGT 183  
 QY 67 PheTyrGlnMetGlyTyrLeuValProGlyLeuMetAlaPheGluAspAsnPheLysAla 86  
 DB 184 TACTGGTGGTACCGTCTCTCCGAGGTTGTATTACGCGCAAGATCTTTCAAGCA 243  
 QY 87 ArgGluThrAspLeuLeuLeuThrThrLeuProLysAlaGlyThrThrThrLysAla 106  
 DB 244 CGACCCAGTCACTCTCCGCTGTAGCTACCAAGACGAGCACCACCTTGGCTCAAGGC 303  
 QY 107 LeuThrPheAlaLeuLeuThrArgAspValAsnHisProSerSerProThrHisProLeu 126  
 DB 304 CTGACTTTCGCAATCGCAATCGTTCCTCGCTTCGATGATTCCTCC-----AACCTCTC 357  
 QY 127 LeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArgGlu 146  
 DB 358 CTGAACGTAACCTCAGAGTTGTTCCTTACATTGAGATAGATTC----- 405  
 QY 147 AsnThrMetProAspLeuAspMetLeuAsnGluSerProArg---LeuPheAlaGlyHis 165  
 DB 406 CTTTCTCTCCCTGAAGTGTGATGTTCTCAAGACAAAGGGAACACTCTGTTTCGATCAT 465  
 QY 166 IleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysLeuLeuAsnIle 185  
 DB 466 ATCCATACGAGTATTACCTGATTCTGTTCTGAATCCGTTGTAAGATGTTTACATA 525  
 QY 186 SerArgAsnArgLysSerThrPheValSerPheTyrLysPheGlyAsnLeuLeuAsnPro 205  
 DB 526 TGGAGAGAACCAAGGACACTTTCATCTCCATGTTGGACTTTC-----CTTCAAG 576  
 QY 206 AspLys-----LeuLeuAspLeuGluLysSerValAspIlePheAlaSer 220  
 DB 577 GAAAGGACAGAGCTTGACCTGTCAGCAATCTTGAGAGTCTTTGATATGTTCTGTCGT 636  
 QY 221 GlyIleSerPheCysGlyProGluTyrAsnPheGlnAlaGluPheThrAsnAlaAlaSer 240  
 DB 637 GGTCTGTCTGGTATGTTCTTATCTTAATCATATCTCGGCTATTCGAAAGCATACCAA 696  
 QY 241 ThrAsn---SerAsnLeuLeuSerTyrGluLysMetLeuGluLysProValGlu 259  
 DB 697 GAGATTCAGATAGATCTTCTTCTCAAGTACGACAGTATGAGAGTGTATCTTTACCG 756  
 QY 260 AsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspGluGluLysGln 279  
 DB 757 TACGTGAAGATCTGGCTGAGTTTATGGTTCATGATTACAGCCGAGGAAGAGAGAAA 816  
 QY 280 GlyLeuValAspGluLeuValLysLeuCysSerPheAsnLeuLysAsnGlnVal 299  
 DB 817 GGTGTGTGTGAAGAGTGTGATCTTTGACGCTTCGAGACGTTGAAGATCTTGAAGCT 876

QY 300 AsnLysAsnGlySerSerTyrAsnSerLys-----IleAspAsnLysHisPhePhe 316  
 DB 877 AACAAAGGGGAAAGACAGAGAGATCGCTCTGTTTACGGAATAGCGGTATTTTC 936  
 QY 317 ArgLysGlyGluValArgAspTyrAlaAsnTyrLeuThrSerGluMetIleLysLeu 336  
 DB 937 AGGAAGGAAGGTGGAGATTGTCGACTATCTGCTCGGAGATGGCTGCTGATA 996  
 QY 337 GluThrAlaGlyLysIleAsnGluSerGluLys 347  
 DB 997 GAT-----GGTTAATGGAAGAGAAATTTAAG 1023  
 RESULT 8  
 AAD54410  
 ID AAD54410 standard; cDNA; 1400 BP.  
 XX  
 AC AAD54410;  
 XX  
 DT 17-JUN-2003 (first entry)  
 XX  
 DE Lolium perenne defensin e (LpDEFe) cDNA.  
 XX  
 KW Ryegrass; thionin; TH; thaumatin-like protein; TL; defensin; DEF; ER;  
 KW elicitor-responsive protein; disease resistance; plant defence response;  
 KW protein storage; pest resistance; genetic marker; gene therapy; antipest;  
 KW agricultural; LpDEFe; gene; ss.  
 XX  
 OS Lolium perenne.  
 XX  
 FH Key Location/Qualifiers  
 CDS 153..833  
 FT /\*tag= a  
 FT /product= "Ryegrass DEFe (LpDEFe) protein"  
 XX  
 PN WO200288359-A1.  
 XX  
 PD 07-NOV-2002.  
 XX  
 PF 01-MAY-2002; 2002WO-AU000539.  
 XX  
 PR 02-MAY-2001; 2001AU-00004735.  
 XX  
 PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 PA (AGRI-) AGRIC VICTORIA SERVICES PTY LTD.  
 XX  
 PI Spangenberg G, Sawbridge TI, Ong EK, Emmerling M;  
 DR WPI: 2003-201227/19.  
 DR P-PSDB; AAE355947.  
 XX  
 PT New ryegrass or fescue nucleic acid encoding a thionin, thaumatin-like,  
 PT elicitor-responsive or defensin polypeptide, for modifying disease and/or  
 PT plant resistance, plant defence response and/or protein storage in a  
 PT plant.  
 XX  
 PS Claim 6; fig 62; 195pp; English.  
 CC The invention relates to novel ryegrass (Lolium) or fescue (Festuca)  
 CC nucleic acids encoding thionin (TH), thaumatin-like (TL), elicitor-  
 CC responsive (ER) or defensin (DEF) proteins and polynucleotides encoding  
 CC such proteins. Sequences of the invention are useful to modify disease  
 CC and/or pest resistance, plant defence response and/or protein storage in  
 CC plants. The nucleotide sequence information and/or single nucleotide  
 CC polymorphisms of the nucleic acid are useful as genetic markers. They are  
 CC also useful in gene therapy. The present sequence is ryegrass DEFe  
 CC (LpDEFe) cDNA  
 XX  
 SQ Sequence 1400 BP; 319 A; 409 C; 380 G; 292 T; 0 U; 0 Other;

Alignment Scores:  
 Pred. No.: 1.13e-44 Length: 1400  
 Score: 558.00 Matches: 150

Percent Similarity: 50.47%		Conservative: 65
Best Local Similarity: 35.21%		Mismatches: 154
Query Match: 27.10%		Indels: 58
DB:		Gaps: 12
US-09-854-122-16 (1-391) x AAD54410 (1-1400)		
QY	1	ThrArgGlyIleThrGlyIleAlaValAlaCysLeuProLeuIleMetAlaGlyIleLeu 20
Db	93	ACGCRAGTAGCGGGAGCTCAGATTAGAGCTAGACACACCTCGCGGA-----146
QY	21	AlaLeuGluLysCysPheGlySerLysAsnGluGlnGluLysGluLysAspSerLysMet 40
Db	147	-----TTTCGCATGGATCCAACTCAACACCAAGACGACGCTCGCAATGCC 191
QY	41	TyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTrpGlyAspThr 60
Db	192	GCAGAGAAAGAGAGCGTCTTGCTCCAGCTCCATGAGGAGGAGATTCTGGAACCG 251
QY	61	MetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAlaPhe 80
Db	252	TACGTCCTTACCAGGCGTCTCGGTATCTCCGAGGTGGCCAGAGCGTCATGCTCTTG 311
QY	81	GluAspAsnPheLysAlaArgGluThrAspIleLeuThrThrLeuProLysAlaGly 100
Db	312	CAAGACGAGTTCAGACCGCGCGCGACGACATCATCTCCGCGCTGCCCGACGCGCTGCCG 371
QY	101	ThrThrTrpThrLysAlaLeuThrPheAlaIleLeu-Thr---ArgAspValAsnHisPr 119
Db	372	ACCACCTGGTCAAGGCGCTTCCCTCCCTCGTCAACCGCTCGCACGCGCGCTGCCG 431
QY	119	oSerSerProThr---HisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeu 138
Db	432	GGAGACGCGGCGGCGCCCGCTGCTTACCACACCGCCCATGACCTCGTCCGCTTCATC 491
QY	139	GluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMet-----154
Db	492	GAG-----AAACCTGACCGGAGCTCTACCTGTCGCGC 524
QY	155	-----LeuAsnGlu-----SerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeu 171
Db	525	GAACCTGGAGGAGCTCGCGTCCCAAGGCTCTCGGCGACCCACATGCCGTTACCGCTGCTG 584
QY	172	ProAlaSerValLeuLysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSer 191
Db	585	CCGGAGAGCATCTCGCGCTCGCTCGCTCGCGTGTATACATCAGCGGAGGCCAAGGAT 644
QY	192	ThrPheValSerPheTrpLysPheGlyAsnLeuIleAsnProAspLysLeuLeuAspLeu 211
Db	645	GTGCTCGTCTCGCTGTCGCACTTGAACCTCGTAAGCAAGACCA-CTTCATCAGCTG 703
QY	212	GluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrp---Asn 230
Db	704	GAGAAAGCCTTCGAGTGTTCAGCAGAGGGGTCTATTTTCGTCAGTGTGGATCAC 763
QY	231	PheGlnAlaGluPheThrAsnAlaIleSerThrAsnSerAsnLeuLeuLeuSerTyr 250
Db	764	TATCTTTGGTACTGGAACAGACGATTCGAGACGCCAGATCGGGTCTCTCTCAAGTAC 823
QY	251	GluGluMetLeuLysProValGluAsnValLysLeuAlaGluPheMetGlyCys 270
Db	824	GAGGATGATGCGCCAGCGGTCAGACAGCTCAAGCCCTGCGCGAGTCTTTCGAGTGC 883
QY	271	GlyPheThrAspAspGluGluLysGlnGlyIleValAspLysLeuValLysLeuCysSer 290
Db	884	CCGTTCCACGTCACGAGGTGGACCGCGAGCTGTGGAGCAGCTCGTGACCTGTGTAGC 943
QY	291	PheAspAsnLeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSer-----308
Db	944	TTTCAGAGCTCAAGAACCTGCTCACTCTCGGGGACATCCAACTCGCTGCTGGTGG 1003
QY	309	---LysIleAspAsnLysHisPhePheArgLysGlyValArgAspTrpAlaSerTyr 327

Db	1004	ATGCCCATGGAGAAGTCCTCTACTTCCAGGAATGGAGCGGTGGCGGATGGGCCAACCCAC 1063
QY	328	LeuThrSerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGluSerGluLys 347
Db	1064	CTGACCCAGAGATGGCCACAGTTGGAT-----GGCATCGTCCAGGAGAGCTCAAA 1117
QY	348	-----HisLeu 349
Db	1118	GGTTCTGCCTCGCCTTTTGGAGATTTCTTCCGATCGATCTGTGAGATTTTCATATT 1177
QY	350	LeuSerAsnLysAsnLeuThrAsnPheAsnLeuAsnAntyrCysGluLysSerAsnIle 369
Db	1178	CTCGTCCCAATATCTCATCTCAATATACTATT---GGATCATGTTATGTTCTACTATC 1234
QY	370	SerLeuCysLeuLeuSerTyrSerPheVal-----IleAsnAsnPheIleLeuLeuLys 387
Db	1235	-----GTTATCTTCTATGGAATAATGTAAGGACCAAGTCGCTTGGTCTATTGCCA 1285
QY	388	LysLysLysLys 391
Db	1286	AAAAAAAAAAAA 1297
RESULT 9		
AAC37521		
ID	AAC37521	standard; DNA; 1205 BP.
XX	AAC37521;	
AC	AAC37521;	
XX	17-OCT-2000	(first entry)
XX	Arabidopsis thaliana	DNA fragment SEQ ID NO: 17688.
XX	Hybridisation assay; genetic mapping; gene expression control;	
XX	protein identification; signal transduction pathway; metabolic pathway;	
XX	promoter; termination sequence; ss.	
XX	Arabidopsis thaliana.	
XX	EP1033405-A2.	
PD	06-SEP-2000.	
XX	25-FEB-2000; 2000EP-00301439.	
XX	25-FEB-1999;	99US-0121825P.
XX	05-MAR-1999;	99US-0123180P.
PR	09-MAR-1999;	99US-0123548P.
PR	23-MAR-1999;	99US-0125788P.
PR	25-MAR-1999;	99US-0126264P.
PR	29-MAR-1999;	99US-0126785P.
PR	01-APR-1999;	99US-0127462P.
PR	06-APR-1999;	99US-0128234P.
PR	08-APR-1999;	99US-0128714P.
PR	16-APR-1999;	99US-0129845P.
PR	19-APR-1999;	99US-0130077P.
PR	21-APR-1999;	99US-0130449P.
PR	23-APR-1999;	99US-0130510P.
PR	23-APR-1999;	99US-0130891P.
PR	28-APR-1999;	99US-0131449P.
PR	30-APR-1999;	99US-0132048P.
PR	30-APR-1999;	99US-0132407P.
PR	04-MAY-1999;	99US-0132484P.
PR	05-MAY-1999;	99US-0132485P.
PR	06-MAY-1999;	99US-0132486P.
PR	06-MAY-1999;	99US-0132487P.
PR	07-MAY-1999;	99US-0132863P.
PR	11-MAY-1999;	99US-0134256P.
PR	14-MAY-1999;	99US-0134218P.
PR	14-MAY-1999;	99US-0134219P.
PR	14-MAY-1999;	99US-0134221P.
PR	14-MAY-1999;	99US-0134370P.
PR	18-MAY-1999;	99US-0134768P.
PR	19-MAY-1999;	99US-0134941P.

us-09-854-122-16.rng

Mon Mar 22 09:47:42 2004

PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139492P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 24-JUN-1999; 99US-0140823P.  
PR 28-JUN-1999; 99US-0140891P.  
PR 29-JUN-1999; 99US-0140891P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 09-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143452P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144086P.  
PR 19-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 19-JUL-1999; 99US-0144335P.  
PR 20-JUL-1999; 99US-0144332P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144844P.  
PR 21-JUL-1999; 99US-0144844P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145132P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145913P.  
PR 27-JUL-1999; 99US-0145918P.  
PR 27-JUL-1999; 99US-0145919P.  
PR 28-JUL-1999; 99US-0145951P.  
PR 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147433P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 09-AUG-1999; 99US-0148171P.  
PR 10-AUG-1999; 99US-0148319P.  
PR 11-AUG-1999; 99US-0148341P.  
PR 12-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 13-AUG-1999; 99US-0149358P.  
PR 16-AUG-1999; 99US-0149317P.  
PR 17-AUG-1999; 99US-0149426P.  
PR 18-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149240P.  
PR 25-AUG-1999; 99US-0150586P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151065P.  
PR 27-AUG-1999; 99US-0151066P.  
PR 27-AUG-1999; 99US-0151080P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157855P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158232P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159329P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 23-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.

PR 26-OCT-1999; 99US-0161361P.  
 PR 28-OCT-1999; 99US-0161920P.  
 PR 28-OCT-1999; 99US-0161932P.  
 PR 28-OCT-1999; 99US-0161933P.  
 PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:  
 Pred. No.: 1,03e-44 Length: 1205  
 Score: 557.50 Matches: 136  
 Percent Similarity: 56.23% Conservative: 67  
 Best Local Similarity: 37.67% Mismatches: 138  
 Query Match: 27.08% Indels: 20  
 DB: 3 Gaps: 9

US-09-854-122-16 (1-391) x AAC37521 (1-1205)

QY 29 LysAsnGluGlnGluLysGluLysGluLysMetTyrLysArgTyrArgGluLeuVal 48  
 Db 64 GAAGAAGAGAAACCAAGTGAAGATTCAAA-----AGTTTGATC 105

QY 49 SerSerLeuProSerAsn---AspTyrTrpGlyAspThrMetArgLeuTyrLysGlyPhe 67  
 Db 106 TCITCACTTCCTTCAGACATAGATTGCTCTGGAGCAAGTTGTACAACTATCAAGATGT 165

QY 68 TrpGlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPhelLysAlaArg 87  
 Db 166 TGTACGATTAAGACATCTCCAGCAATCTCCAAATTCACAAATCTTTCAGCCACAA 225

QY 88 GluThrAspIleLeuLeuThrThrLeuProLysAlaGlyThrThrTrpThrLysAlaLeu 107  
 Db 226 GAAACCGATATATGTTGCTCTTCCCAATCAATCAGTACGCTGCTCAAGGCATC 285

QY 108 ThrPheAlaLeuLeuThrArgAspValAsnHisProSerSerProThrHisProLeuLeu 127  
 Db 286 ACATTCGCATCTCGCTCAAGATCAAAACAT-----ACTTCAGAAATCATCTCTGCTA 339

QY 128 PhePheAsnProHisSerCysValGlnAsnLeuGluTyr---LeuTyrMetGlyArgGlu 146  
 Db 340 ACTCATATCTCATGAGCTAGTGGCTACCTCGACTCGATCTTATCTC-----AAA 393

QY 147 AsnThrMetProAspLeuAspMetLeu---AsnGluSerProArgLeuPheAlaGlyHis 165  
 Db 394 AGTCGAAACCGGATATGTCCAAAGTACCATCATCTCCGAGATTTGTCTCAACCCAC 453

QY 166 IleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIleLeuAsnIle 185  
 Db 454 ATGCTTTTCGATCGCTTAAAGTACCATCAAGGACACTCTTCGACAGATAGTATGTG 513

QY 186 SerArgAsnArgLysSerThrPheValSerPheTrpLysPheGlyAsnLeuLeuAsnPro 205  
 Db 514 TGCAGGAACGTAAAGACGTTGGTGTATCACTTTGTTGTTTCGAAACTCCATTAGTGA 573

QY 206 AspLysLeuLeuAspLeuGluLysSerValAspIlePheAlaSerGlyLysSerPheCys 225  
 Db 574 GAAACAAATTAAGTCTCGAGGCTTTGTTTCGAGTCTTTATGTCGGAGTTAACTATGC 633

QY 226 GlyProGluTrp---AsnPheGlnAlaGluPheThrAsnAlaAlaSerThrAsnSerAsn 244  
 Db 634 GGTCCTTGTGGAAATATGTTAGGCTATTGAGAGGAAGCTTGGAGATCTTAAGCAT 693

QY 245 LeuLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnValLysLysLeu 264  
 Db 694 GTGCTTTCTTGTAGGTACGAGGAGTTGAAGACGAGCTCTGTCGAAATCAAGACTT 753

QY 265 AlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGlyGlyLeuValAspGlu 284  
 Db 754 GCAGAGTCTTAGATTTTCCATTCACAAAGGAAGAGAGATGAGGAGGTGTAGACAA 813

QY 285 IleValLysLeuCysSerPheAspAsnLeuLysAsnGlnGlnValAsnLysAsnGlySer 304  
 Db 814 ATCTTGAACTTTGTTCTCTAAGAAACCTTAGCGGTTTCGAGATCAACAAACACAGG 870

QY 305 SerTyrAsnSerLysIleAspAsnLysHisPheArgLysGlyGluValArgAspTrp 324

Db 871 AGCTTGTCCGAAGCAGTAAGTTTCAAGAGATTTTCCGTAAGGGAAGTGGTGATTGG 930  
 QY 325 AlaAsnTyrLeuThrSerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGlu 344  
 Db 931 AAGAGTTATATGCTCTCTGAAATGGAAACAAATCGAC-----ATGATGTGT 978

QY 345 SerGluLysHisLeuLeuSerAsnLysAsnLeuThrAsnPheAsnLeuAsnTyrCys 364  
 Db 979 GAGGAGAAACTTCAAGGCTCTGGTTTGATTTAGATTTAGATCTCATCTATGTATGTG 1038

QY 365 GluLysSerAsnIle-SerLeuCysLeuLeuSerTyrSerPheValIleAsnAsnPheIle 384  
 Db 1039 AACAGGTTTAAATCTCAACCTTAATATGCTGGTTGTCCTTTTCTTGTATCAATGTAAT 1098

QY 384 e 384  
 Db 1099 A 1099

RESULT 10  
 AAF29177  
 ID AAF29177 standard; DNA; 1077 BP.  
 XX  
 AC AAF29177;  
 XX  
 DT 09-APR-2001 (first entry)  
 XX  
 DE Hydroxyjasmonic acid sulfoltransferase AtST2a gene.  
 XX  
 KW Hydroxyjasmonic acid sulfoltransferase; AtST2a; flowering time; cabbage;  
 KW jasmonate; genetically modified plant; lettuce; sugar cane; carrot;  
 KW increase vegetative growth; biomass increase; ds.  
 XX  
 OS Arabidopsis thaliana.  
 XX  
 PN WO200102589-A2.  
 XX  
 PD 11-JAN-2001.  
 XX  
 PF 06-JUL-2000; 2000WO-CA000801.  
 XX  
 PR 06-JUL-1999; 99CA-02274873.  
 XX  
 PA (VARI/) VARIN L.  
 PA (GIDD/) GIDDA S.  
 XX  
 PI Varin L, Gidda S;  
 XX  
 DR WPI; 2001-159272/16.  
 XX  
 DR P-PSDB; AAB49722.  
 XX

Methods for modulating flowering in plants, particularly useful for plants used in the food-processing industry, involves modifying the endogenous level of compounds of the jasmonate family.

Claim 38; Fig 7; 50pp; English.

This invention relates to a method for modulating flowering in a plant. The method comprises modifying the endogenous level of at least one compound of the jasmonate family in a plant. The methods are used to produce plants which are genetically modified to flower early or tardily when compared to a corresponding plant that is not genetically modified, where the modified plant has an increased (flower early) or lowered (flower tardily) level of jasmonic acid or a compound of the jasmonate family. The method is useful for modulating flowering, particularly for plants that are used in the food-processing industry and plants with horticultural value. The method is particularly useful for e.g. delaying flowering time in crops like lettuce, cabbage, sugar cane or carrots, which results in increased vegetative growth and biomass. The present sequence represents the Arabidopsis thaliana AtST2a gene, which encodes an hydroxyjasmonic acid sulfoltransferase protein, which can be used in the method of the invention

SQ Sequence 1077 BP; 294 A; 257 C; 254 G; 272 T; 0 U; 0 Other;

Alignment Scores:  
Pred. No.: 1,4e-44 Length: 1077  
Score: 555.50 Matches: 123  
Percent Similarity: 55.98% Conservative: 69  
Best Local Similarity: 35.86% Mismatches: 140  
Query Match: 26.98% Indels: 11  
DB: 4 Gaps: 8

US-09-854-122-16 (1-391) x AAF29177 (1-1077)

QY 1 ThrArgGlyIleThrGlyIleAlaValAlaCysLeuProLeuIleMetAlaGlyIleIeu 20  
DB 7 ACCTCAAGCATGAAGAGCATTCATGCGCATCCCAAGTTCTCTCATGTGCACAGCTC 66  
QY 21 AlaLeuGluLysCysPheGlySerLysAsnGluGlnGlyLysGluGluAspSerLysMet 40  
DB 67 GAGCTCTTAAAGAGGCAAACTCGCAGCTCCCGAAAGCGGAAGATGAAGGGCTA 126  
QY 41 TyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTyrGlyAspThr 60  
DB 127 AGCTCGAGTTTCCCAAGATGTGTGATTTCTTCTTAAGGAGAGAGGATGGAGAACTCGT 186  
QY 61 ---MetArgLeuTyrLysGlyPheTyrGlnMetGlyTyrLeuValProGlyIleMetAla 79  
DB 187 TACCTTTACCTATTCCAGGGTTTGGTCCCAAGCAAGAGATTCAAGCCCATGTGCT 246  
QY 80 PheGluAspAsnPhelLysAlaArgGluThrAspIleLeuThrLeuProLysAla 99  
DB 247 TTCCAAAACATTTCCAACTCCCTCGAAAACGACGTCGTTCGCCACCATCTAAATGCC 306  
QY 100 GlyThrThrThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisPro 119  
DB 307 GGTACAACTGCTGCTTAAGCTTTAATCTTACCATCTTAAACCGTCACCGTTTATCCG 366  
QY 120 -----SerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsn 137  
DB 367 GTTGCCTCGATTACCAACACCTCTTTTCACTTCACACCTCATGACCTTGATCTTC 426  
QY 138 LeuGluTyr---LeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn 156  
DB 427 TTCGATGACAAGCTTTACGCCCAACGGAGAT-----GTTCGCCGATCTCTCGGGTCTA--- 477  
QY 157 GluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeu 176  
DB 478 GCCAGTCCAGACGTTCCGACCCCACTTACGTTCCGTTCCCTAAAGGAAACGATCGAG 537  
QY 177 LysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheValSerPhe 196  
DB 538 AAACCCGGTGTGAAGTCTGTACTTGTGCCGGAACCCGTTTGCACACATCACTCTTCG 597  
QY 197 TrpLysPheGlyAsnLeuIleAsnProAspLysLeu-----LeuAspLeuGluLysSer 214  
DB 598 TGGCAITACACCAACCAATCAATCCGAGTCAGTCAGGCCCACTCTGTAGACCAAGCT 657  
QY 215 ValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGlu 234  
DB 658 TTTGATCTGTATTGCCGGGAGTATCGGGTTTGGCCCGCTTTTGGGAACACATGTGGGA 717  
QY 235 Phe---ThrAsnAlaIleSerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMet 253  
DB 718 TACTGGAGAGAGCTTGAAGACACAGAGAAAGTCTTTTAAAGTACGAGGATCTC 777  
QY 254 LeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThr 273  
DB 778 AAAGACACATCGAGACCACTTGAAGAGCTTGCACCTTCTTAGAGCTTCTTTTACC 837  
QY 274 AspAspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPheAspAsn 293  
DB 838 GAAGAAGGAGCAACAAAGGAGTGTGAAGCTATCCCGAGCTGTGTAGCTTCGAGAT 897  
QY 294 LeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLys 313

DB 898 CTGAAGAAGTTCGAGTGAACAAAGTCAACAAGTCATCAAGAAC---TTTGAGATCGA 954  
QY 314 HisPhePheArgLysGlyIleValArgAspTrpAlaAsnTyrLeuThrSerGluMetIle 333  
DB 955 TTCTTGTTCGGAAGAGAGAGTGAAGTGGTAACTATTTGTACCTTACCTTCACAAGTG 1014  
QY 334 LysLysLeu 336  
DB 1015 GAAAGATTG 1023  
RESULT 11  
AAC48839  
ID AAC48839 standard; DNA; 1270 BP.  
XX  
AC AAC48839;  
XX  
DT 18-OCT-2000 (first entry)  
XX  
DE Arabidopsis thaliana DNA fragment SEQ ID NO: 58964.  
XX  
KW Hybridisation assay; Genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway; metabolic pathway;  
KW promoter; termination sequence; ss.  
OS Arabidopsis thaliana.  
XX  
PN EP1033405-A2.  
XX  
PD 06-SEP-2000.  
XX  
PF 25-FEB-2000; 2000EP-00301439.  
XX  
PR 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 28-APR-1999; 99US-0130891P.  
PR 30-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 07-MAY-1999; 99US-0132487P.  
PR 11-MAY-1999; 99US-0132863P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 18-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134788P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136382P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.



PR 10-JUN-1999; 99US-0138540P.  
 PR 10-JUN-1999; 99US-0138847P.  
 PR 14-JUN-1999; 99US-0139119P.  
 PR 16-JUN-1999; 99US-0139452P.  
 PR 16-JUN-1999; 99US-0139453P.  
 PR 17-JUN-1999; 99US-0139454P.  
 PR 18-JUN-1999; 99US-0139455P.  
 PR 18-JUN-1999; 99US-0139456P.  
 PR 18-JUN-1999; 99US-0139457P.  
 PR 18-JUN-1999; 99US-0139458P.  
 PR 18-JUN-1999; 99US-0139459P.  
 PR 18-JUN-1999; 99US-0139460P.  
 PR 18-JUN-1999; 99US-0139461P.  
 PR 18-JUN-1999; 99US-0139462P.  
 PR 18-JUN-1999; 99US-0139463P.  
 PR 18-JUN-1999; 99US-0139750P.  
 PR 18-JUN-1999; 99US-0139751P.  
 PR 21-JUN-1999; 99US-0139817P.  
 PR 22-JUN-1999; 99US-0139899P.  
 PR 23-JUN-1999; 99US-0140353P.  
 PR 23-JUN-1999; 99US-0140354P.  
 PR 24-JUN-1999; 99US-0140695P.  
 PR 28-JUN-1999; 99US-0140823P.  
 PR 29-JUN-1999; 99US-0140991P.  
 PR 30-JUN-1999; 99US-0141287P.  
 PR 01-JUL-1999; 99US-0141842P.  
 PR 01-JUL-1999; 99US-0142154P.  
 PR 02-JUL-1999; 99US-0142055P.  
 PR 06-JUL-1999; 99US-0142390P.  
 PR 08-JUL-1999; 99US-0142803P.  
 PR 09-JUL-1999; 99US-0142920P.  
 PR 12-JUL-1999; 99US-0142977P.  
 PR 13-JUL-1999; 99US-0143542P.  
 PR 14-JUL-1999; 99US-0143624P.  
 PR 15-JUL-1999; 99US-0144008P.  
 PR 16-JUL-1999; 99US-0144088P.  
 PR 16-JUL-1999; 99US-0144089P.  
 PR 19-JUL-1999; 99US-0144325P.  
 PR 19-JUL-1999; 99US-0144331P.  
 PR 19-JUL-1999; 99US-0144332P.  
 PR 19-JUL-1999; 99US-0144333P.  
 PR 19-JUL-1999; 99US-0144334P.  
 PR 19-JUL-1999; 99US-0144335P.  
 PR 20-JUL-1999; 99US-0144352P.  
 PR 20-JUL-1999; 99US-0144632P.  
 PR 20-JUL-1999; 99US-0144884P.  
 PR 21-JUL-1999; 99US-0144814P.  
 PR 21-JUL-1999; 99US-0145086P.  
 PR 21-JUL-1999; 99US-0145088P.  
 PR 22-JUL-1999; 99US-0145085P.  
 PR 22-JUL-1999; 99US-0145087P.  
 PR 22-JUL-1999; 99US-0145089P.  
 PR 22-JUL-1999; 99US-0145192P.  
 PR 22-JUL-1999; 99US-0145145P.  
 PR 23-JUL-1999; 99US-0145218P.  
 PR 23-JUL-1999; 99US-0145224P.  
 PR 26-JUL-1999; 99US-0145276P.  
 PR 27-JUL-1999; 99US-0145913P.  
 PR 27-JUL-1999; 99US-0145918P.  
 PR 27-JUL-1999; 99US-0145919P.  
 PR 28-JUL-1999; 99US-0145951P.  
 PR 02-AUG-1999; 99US-0146386P.  
 PR 02-AUG-1999; 99US-0146388P.  
 PR 03-AUG-1999; 99US-0146389P.  
 PR 03-AUG-1999; 99US-0147038P.  
 PR 04-AUG-1999; 99US-0147204P.  
 PR 04-AUG-1999; 99US-0147302P.  
 PR 05-AUG-1999; 99US-0147192P.  
 PR 05-AUG-1999; 99US-0147260P.  
 PR 06-AUG-1999; 99US-0147303P.  
 PR 06-AUG-1999; 99US-0147416P.  
 PR 09-AUG-1999; 99US-0147493P.

PR 09-AUG-1999; 99US-0147935P.  
 PR 10-AUG-1999; 99US-0148171P.  
 PR 11-AUG-1999; 99US-0148319P.  
 PR 12-AUG-1999; 99US-0148341P.  
 PR 13-AUG-1999; 99US-0148565P.  
 PR 13-AUG-1999; 99US-0148684P.  
 PR 16-AUG-1999; 99US-0149368P.  
 PR 17-AUG-1999; 99US-0149175P.  
 PR 18-AUG-1999; 99US-0149426P.  
 PR 18-AUG-1999; 99US-0149722P.  
 PR 20-AUG-1999; 99US-0149723P.  
 PR 20-AUG-1999; 99US-0149829P.  
 PR 23-AUG-1999; 99US-0149902P.  
 PR 23-AUG-1999; 99US-0149930P.  
 PR 25-AUG-1999; 99US-0150566P.  
 PR 26-AUG-1999; 99US-0150884P.  
 PR 27-AUG-1999; 99US-0151065P.  
 PR 27-AUG-1999; 99US-0151066P.  
 PR 27-AUG-1999; 99US-0151080P.  
 PR 30-AUG-1999; 99US-0151503P.  
 PR 31-AUG-1999; 99US-0151438P.  
 PR 01-SEP-1999; 99US-0151530P.  
 PR 07-SEP-1999; 99US-0152363P.  
 PR 10-SEP-1999; 99US-0153070P.  
 PR 13-SEP-1999; 99US-0153758P.  
 PR 15-SEP-1999; 99US-0154018P.  
 PR 16-SEP-1999; 99US-0154039P.  
 PR 20-SEP-1999; 99US-0154179P.  
 PR 22-SEP-1999; 99US-0155139P.  
 PR 23-SEP-1999; 99US-0155486P.  
 PR 24-SEP-1999; 99US-0155559P.  
 PR 28-SEP-1999; 99US-0156458P.  
 PR 29-SEP-1999; 99US-0156596P.  
 PR 04-OCT-1999; 99US-0157117P.  
 PR 05-OCT-1999; 99US-0157753P.  
 PR 06-OCT-1999; 99US-0157855P.  
 PR 07-OCT-1999; 99US-0158029P.  
 PR 08-OCT-1999; 99US-0158232P.  
 PR 12-OCT-1999; 99US-0158369P.  
 PR 13-OCT-1999; 99US-0159293P.  
 PR 13-OCT-1999; 99US-0159294P.  
 PR 13-OCT-1999; 99US-0159295P.  
 PR 14-OCT-1999; 99US-0159329P.  
 PR 14-OCT-1999; 99US-0159330P.  
 PR 14-OCT-1999; 99US-0159331P.  
 PR 14-OCT-1999; 99US-0159637P.  
 PR 14-OCT-1999; 99US-0159638P.  
 PR 18-OCT-1999; 99US-0159584P.  
 PR 21-OCT-1999; 99US-0160741P.  
 PR 21-OCT-1999; 99US-0160767P.  
 PR 21-OCT-1999; 99US-0160768P.  
 PR 21-OCT-1999; 99US-0160770P.  
 PR 21-OCT-1999; 99US-0160814P.  
 PR 21-OCT-1999; 99US-0160815P.  
 PR 22-OCT-1999; 99US-0160800P.  
 PR 22-OCT-1999; 99US-0160981P.  
 PR 22-OCT-1999; 99US-0160989P.  
 PR 25-OCT-1999; 99US-0161404P.  
 PR 25-OCT-1999; 99US-0161405P.  
 PR 25-OCT-1999; 99US-0161406P.  
 PR 26-OCT-1999; 99US-0161359P.  
 PR 26-OCT-1999; 99US-0161360P.  
 PR 26-OCT-1999; 99US-0161361P.  
 PR 26-OCT-1999; 99US-0161940P.  
 PR 28-OCT-1999; 99US-0161922P.  
 PR 28-OCT-1999; 99US-0161993P.  
 PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:

Pred. No.: 1.74e-44

Score: 555.50

Percent Similarity: 55.98%

Best Local Similarity: 35.86%

Length: 1270

Matches: 123

Conservative: 69

Mismatches: 140

Query Match:	26.98%	Indels:	11
DB:	3	Gaps:	8
US-09-854-122-16 (1-391) x AAC48839 (1-1270)			
Qy	1	ThrArgGlyIleThrGlyIleAlaValAlaCysLeuProLeuIleMetAlaGlyIleLeu	20
Db	70	ACCTCAAGCATGAAGAGCATTCCTCAATGGCGATCCCAAGTTTCTCCATGTGTCAACAAGCTC	129
Qy	21	AlaLeuGluLysCysPheGlySerLysAsnGluGlnGluLysGluGluAspSerLysMet	40
Db	130	GAGTCTCTTAAGAGGCAAACTCCGACGTCCTCGAAGCCGAGAGAGATGAAGGGCTA	189
Qy	41	TyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTyrGlyAspThr	60
Db	190	AGTGGAGTTCGAAGAGATGTGGATTCTCTTCTTAAGGAGAGAGATGGAGAACTCGT	249
Qy	61	MetArgLeuTyrLysGlyPheTyrGlnMetGlyTyrLeuValProGlyIleMetAla	79
Db	250	TACCTTTTACCTATTCGAAGGGTTTGTGTCGAAGCCCAAGAGATTCAGCCATCATGTCT	309
Qy	80	PheGluAspAsnPhelysAlaArgGluThrAspIleIleLeuThrThrLeuProLysAla	99
Db	310	TTCCAAACAAATTCCTCAATCCCTCGAAACAGCGTCTTCGCCACCATACCTAAATCC	369
Qy	100	GlyThrThrThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisPro	119
Db	370	GGTACAACTGGCTAAAGCTTTTAACTTCACCTTCCACCTCATGACCTTGACCTTTC	429
Qy	120	-----SerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsn	137
Db	430	GTTGCTCGAGTACCAACCACTCTTTTCACTTCCACCTCATGACCTTGACCTTTC	489
Qy	138	LeuGluTyr---LeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn	156
Db	490	TTGAGTACAGCTTTACGCCAACCGAGAT-----GTTCCCGATCTCCGGGTCTA---	540
Qy	157	GluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeu	176
Db	541	GCAGTCCAGACAGCTTGGCAACCACTTACCGTTCGGTTCCTTAAGGAACGATCGAG	600
Qy	177	LysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheValSerPhe	196
Db	601	AAACCCGGTGTGAAGTCTGTGTTACTTGTGCGGACCCGTTTGACACATTCATCTCTCG	660
Qy	197	TrpLysPheGlyAsnLeuIleAsnProAspLysLeu-----LeuAspLeuGluLysSer	214
Db	661	TGGCATTACCAACCACTCAATCCGATCGATCGAGCCGCTGTGTAGACCAAGCT	720
Qy	215	ValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGlu	234
Db	721	TTTGATCTGTATTCGGGGAGTGTGCGGTGTGGCCGTTTGGGAACACATGTTGGGA	780
Qy	235	Phe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMet	253
Db	781	TACTGGAGAGAGCTTGAAGACGACGAGAAAGTCTCTTTTAAGTACGAGGATCTC	840
Qy	254	LeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThr	273
Db	841	AAAGACGACATCGAGACCACTTGAAGAGGCTTTCGCACTTCTTAGAGCTTCTTCACC	900
Qy	274	AspAspGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPheAspAsn	293
Db	901	GAAGAAGAGAGCAAGAGGAGTGTGAAGGCTATCCCGAGCTGTGTAGCTTCGAGAT	960
Qy	294	LeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLys	313
Db	961	CTGAGAGAGTGGAGGTGAACAGTCAACAGTCGATCAAGAAC---TTTGAGATCGA	1017
Qy	314	HisPhePheArgLysGlyGluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIle	333
Db	1018	TTCTTGTTCGGAAGAGAGAGTGAAGTGAATGGGTAACTATTGTGTCACCTTCACAGTG	1077

Qy 334 LysLysLeu 336  
Db 1078 GAAAGATTG 1086

RESULT 12  
AAC47848  
ID AAC47848 standard; DNA; 1273 BP.  
XX AC AAC47848;  
XX 18-OCT-2000 (first entry)  
XX Arabidopsis thaliana DNA fragment SEQ ID NO: 55342.  
XX Hybridisation assay; genetic mapping; gene expression control;  
XX protein identification; signal transduction pathway; metabolic pathway;  
XX promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
XX EPI033405-A2.  
XX 06-SEP-2000.  
XX 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
XX 05-MAR-1999; 99US-0123180P.  
XX 09-MAR-1999; 99US-0123548P.  
XX 23-MAR-1999; 99US-0125788P.  
XX 25-MAR-1999; 99US-0126264P.  
XX 29-MAR-1999; 99US-0126785P.  
XX 01-APR-1999; 99US-0127462P.  
XX 06-APR-1999; 99US-0128234P.  
XX 08-APR-1999; 99US-0128714P.  
XX 16-APR-1999; 99US-0129845P.  
XX 19-APR-1999; 99US-0130077P.  
XX 21-APR-1999; 99US-0130449P.  
XX 23-APR-1999; 99US-0130510P.  
XX 28-APR-1999; 99US-0130891P.  
XX 30-APR-1999; 99US-0131449P.  
XX 30-APR-1999; 99US-0132048P.  
XX 30-APR-1999; 99US-0132407P.  
XX 04-MAY-1999; 99US-0132484P.  
XX 05-MAY-1999; 99US-0132485P.  
XX 06-MAY-1999; 99US-0132486P.  
XX 06-MAY-1999; 99US-0132487P.  
XX 07-MAY-1999; 99US-0132863P.  
XX 11-MAY-1999; 99US-0134256P.  
XX 14-MAY-1999; 99US-0134218P.  
XX 14-MAY-1999; 99US-0134219P.  
XX 14-MAY-1999; 99US-0134221P.  
XX 14-MAY-1999; 99US-0134370P.  
XX 18-MAY-1999; 99US-0134768P.  
XX 19-MAY-1999; 99US-0134941P.  
XX 20-MAY-1999; 99US-0135124P.  
XX 21-MAY-1999; 99US-0135353P.  
XX 24-MAY-1999; 99US-0135629P.  
XX 25-MAY-1999; 99US-0136021P.  
XX 27-MAY-1999; 99US-0136332P.  
XX 28-MAY-1999; 99US-0136782P.  
XX 01-JUN-1999; 99US-0137222P.  
XX 03-JUN-1999; 99US-0137528P.  
XX 10-JUN-1999; 99US-0138540P.  
XX 10-JUN-1999; 99US-0138847P.  
XX 14-JUN-1999; 99US-0139119P.  
XX 16-JUN-1999; 99US-0139452P.  
XX 16-JUN-1999; 99US-0139453P.  
XX 17-JUN-1999; 99US-0139492P.  
XX 18-JUN-1999; 99US-0139454P.

PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.  
PR 18-JUN-1999; 99US-0139462P.  
PR 18-JUN-1999; 99US-0139463P.  
PR 18-JUN-1999; 99US-0139750P.  
PR 18-JUN-1999; 99US-0139763P.  
PR 21-JUN-1999; 99US-0139817P.  
PR 22-JUN-1999; 99US-0139899P.  
PR 23-JUN-1999; 99US-0140353P.  
PR 23-JUN-1999; 99US-0140354P.  
PR 24-JUN-1999; 99US-0140695P.  
PR 24-JUN-1999; 99US-0140823P.  
PR 28-JUN-1999; 99US-0140991P.  
PR 30-JUN-1999; 99US-0141287P.  
PR 01-JUL-1999; 99US-0141842P.  
PR 01-JUL-1999; 99US-0142154P.  
PR 02-JUL-1999; 99US-0142055P.  
PR 06-JUL-1999; 99US-0142390P.  
PR 08-JUL-1999; 99US-0142803P.  
PR 08-JUL-1999; 99US-0142920P.  
PR 12-JUL-1999; 99US-0142977P.  
PR 13-JUL-1999; 99US-0143542P.  
PR 14-JUL-1999; 99US-0143624P.  
PR 15-JUL-1999; 99US-0144005P.  
PR 16-JUL-1999; 99US-0144085P.  
PR 16-JUL-1999; 99US-0144325P.  
PR 19-JUL-1999; 99US-0144331P.  
PR 19-JUL-1999; 99US-0144332P.  
PR 19-JUL-1999; 99US-0144333P.  
PR 19-JUL-1999; 99US-0144334P.  
PR 20-JUL-1999; 99US-0144352P.  
PR 20-JUL-1999; 99US-0144632P.  
PR 20-JUL-1999; 99US-0144884P.  
PR 21-JUL-1999; 99US-0144814P.  
PR 21-JUL-1999; 99US-0145086P.  
PR 21-JUL-1999; 99US-0145088P.  
PR 22-JUL-1999; 99US-0145085P.  
PR 22-JUL-1999; 99US-0145087P.  
PR 22-JUL-1999; 99US-0145089P.  
PR 22-JUL-1999; 99US-0145192P.  
PR 23-JUL-1999; 99US-0145145P.  
PR 23-JUL-1999; 99US-0145218P.  
PR 23-JUL-1999; 99US-0145224P.  
PR 26-JUL-1999; 99US-0145276P.  
PR 27-JUL-1999; 99US-0145313P.  
PR 27-JUL-1999; 99US-0145318P.  
PR 27-JUL-1999; 99US-0145319P.  
PR 28-JUL-1999; 99US-0145351P.  
PR 02-AUG-1999; 99US-0146386P.  
PR 02-AUG-1999; 99US-0146388P.  
PR 02-AUG-1999; 99US-0146389P.  
PR 03-AUG-1999; 99US-0147038P.  
PR 04-AUG-1999; 99US-0147204P.  
PR 04-AUG-1999; 99US-0147302P.  
PR 05-AUG-1999; 99US-0147192P.  
PR 05-AUG-1999; 99US-0147260P.  
PR 06-AUG-1999; 99US-0147303P.  
PR 06-AUG-1999; 99US-0147416P.  
PR 09-AUG-1999; 99US-0147493P.  
PR 09-AUG-1999; 99US-0147935P.  
PR 10-AUG-1999; 99US-0148171P.  
PR 11-AUG-1999; 99US-0148199P.  
PR 12-AUG-1999; 99US-0148341P.  
PR 13-AUG-1999; 99US-0148565P.  
PR 13-AUG-1999; 99US-0148684P.  
PR 16-AUG-1999; 99US-0149368P.

PR 17-AUG-1999; 99US-0149175P.  
PR 18-AUG-1999; 99US-0149426P.  
PR 20-AUG-1999; 99US-0149722P.  
PR 20-AUG-1999; 99US-0149723P.  
PR 20-AUG-1999; 99US-0149929P.  
PR 23-AUG-1999; 99US-0149902P.  
PR 23-AUG-1999; 99US-0149930P.  
PR 25-AUG-1999; 99US-0150566P.  
PR 26-AUG-1999; 99US-0150884P.  
PR 27-AUG-1999; 99US-0151086P.  
PR 27-AUG-1999; 99US-0151086P.  
PR 27-AUG-1999; 99US-0151086P.  
PR 30-AUG-1999; 99US-0151303P.  
PR 31-AUG-1999; 99US-0151438P.  
PR 01-SEP-1999; 99US-0151930P.  
PR 07-SEP-1999; 99US-0152363P.  
PR 10-SEP-1999; 99US-0153070P.  
PR 13-SEP-1999; 99US-0153758P.  
PR 15-SEP-1999; 99US-0154018P.  
PR 16-SEP-1999; 99US-0154039P.  
PR 20-SEP-1999; 99US-0154779P.  
PR 22-SEP-1999; 99US-0155139P.  
PR 23-SEP-1999; 99US-0155486P.  
PR 24-SEP-1999; 99US-0155659P.  
PR 28-SEP-1999; 99US-0156458P.  
PR 29-SEP-1999; 99US-0156596P.  
PR 04-OCT-1999; 99US-0157117P.  
PR 05-OCT-1999; 99US-0157753P.  
PR 06-OCT-1999; 99US-0157865P.  
PR 07-OCT-1999; 99US-0158029P.  
PR 08-OCT-1999; 99US-0158222P.  
PR 12-OCT-1999; 99US-0158369P.  
PR 13-OCT-1999; 99US-0159293P.  
PR 13-OCT-1999; 99US-0159294P.  
PR 13-OCT-1999; 99US-0159295P.  
PR 14-OCT-1999; 99US-0159330P.  
PR 14-OCT-1999; 99US-0159331P.  
PR 14-OCT-1999; 99US-0159637P.  
PR 14-OCT-1999; 99US-0159638P.  
PR 18-OCT-1999; 99US-0159584P.  
PR 21-OCT-1999; 99US-0160741P.  
PR 21-OCT-1999; 99US-0160767P.  
PR 21-OCT-1999; 99US-0160768P.  
PR 21-OCT-1999; 99US-0160770P.  
PR 21-OCT-1999; 99US-0160814P.  
PR 21-OCT-1999; 99US-0160815P.  
PR 22-OCT-1999; 99US-0160980P.  
PR 22-OCT-1999; 99US-0160981P.  
PR 22-OCT-1999; 99US-0160989P.  
PR 25-OCT-1999; 99US-0161404P.  
PR 25-OCT-1999; 99US-0161405P.  
PR 25-OCT-1999; 99US-0161406P.  
PR 26-OCT-1999; 99US-0161359P.  
PR 26-OCT-1999; 99US-0161360P.  
PR 26-OCT-1999; 99US-0161361P.  
PR 28-OCT-1999; 99US-0161920P.  
PR 28-OCT-1999; 99US-0161992P.  
PR 28-OCT-1999; 99US-0161993P.  
PR 29-OCT-1999; 99US-0162142P.

Alignment Scores:  
Pred. No.: 1.74e-44  
Score: 555.50  
Percent Similarity: 55.98%  
Best Local Similarity: 35.86%  
Query Match: 26.98%  
DB: 3

Length: 1273  
Matches: 123  
Conservative: 69  
Mismatch: 140  
Indels: 11  
Gaps: 8

US-09-854-122-16 (1-391) x AAC47848 (1-1273)

QY 1 ThrArgGlyIleThrGlyIleAlaValAlaCysLeuProleuIleMetAlaGlyIleLeu 20  
||| :|||  
||| :|||

Mon Mar 22 09:47:42 2004

us-09-854-122-16.rng

Db 72 ACCTCAGCAGTGAAGAGCATTCCTCAATGCGGATCCCAAGTTCTCCATGTCACAGGTC 131  
Qy 21 AlaLeuGluLysCysPheGlySerLysAsnGluGlnGluLysGluAspSerLysMet 40  
Db 132 GAGCTCTCTAAAGAGGCAAACTCGCAGCTCCGAAAGCCGAGAGATGAAGGGCTA 191  
Qy 41 TyrLysArgTyrArgGluLeuValSerSerLysLeuProSerAsnAspTyrTyrGlyAspThr 60  
Db 192 ACCTCGGAGTTCGAAGAGATGTGGATCTCTTCCTTAGGAGAGGATGAGAACTCGT 251  
Qy 61 ---MetArgLeuTyrLysGlyPheTyrGlnMetGlyTyrLeuValProGlyLeuMetAla 79  
Db 252 TACCTTTACCTATTCAAGGGTTTGGTGCCAAAGCAAGAGATTCAAGCCCATCATGTCT 311  
Qy 80 PheGluAspAsnPhelLysAlaArgGluThrAspLeuLeuThrAspLeuProLysAla 99  
Db 312 TTCCAAAACATTTCAATCCCTCGAAAGACGCTGCTCGCCACCATCACTAAATCC 371  
Qy 100 GlyThrThrTyrLysAlaLeuThrPheAlaLeuLeuThrArgAspValAsnHisPro 119  
Db 372 GGTACAACTGGCTAAAGCTTTAACTTTCAACCATCTTAACCGTCACCGGTTTGATCCG 431  
Qy 120 -----SerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsn 137  
Db 432 GTTGCTCGAGTACCAACACCTCTTTCACTTCCACCTCATGACCTTGATCTTTC 491  
Qy 138 LeuGluTyr---LeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn 156  
Db 492 TTCGAGTACAGCTTTACGCCAACCGAGAT-----GTTCGCCGATCTCTCGGGTCTA--- 542  
Qy 157 GluSerProArgLeuPheAlaGlyHisLeuProTyrSerLeuLeuProAlaSerValLeu 176  
Db 543 GCCAGTCCAGAACGTTCCCAACCCCTTACCGTTCCTTAAGGAAACGATCGAG 602  
Qy 177 LysSerGlyThrLysLeuLeuSerArgAsnArgLysSerThrPheValSerPhe 196  
Db 603 AAACCCGGTGAAGTGGTGTACTGTGCCGAACCCGTTGACACATTCATCTCTTCG 662  
Qy 197 TrpLysPheGlyAsnLeuLeuAsnProAspLysLeu-----LeuAspLeuGluLysSer 214  
Db 663 TGGCATTTACACCAACACATCAATCCGAGTCAGTGAGCCAGCTCTTGCTAGACCAAGCT 722  
Qy 215 ValAspPheAlaSerGlyLeuSerPheCysGlyProGluTrpAsnPheGlnAlaGlu 234  
Db 723 TTGTATCTGTATTCGCGGAGTGATCGGGTTTGGCCCGTTTGGGAACACATGTTGGGA 782  
Qy 235 Phe---ThrAsnAlaLaserThrAsnSerAsnLeuLeuLeuSerTyrGluGluMet 253  
Db 783 TACTGGAGAGAGAGCTTGAAGAGACCAAGAGAAAGTCTCTTTTAAAGTACGAGGATCTC 842  
Qy 254 LeuGluLysProValGluAsnValLysLeuAlaGluPheMetGlyCysGlyPheThr 273  
Db 843 AAAGACGACATCGAGACCACTTGAAGAGCTTGCAACTTCTTAGAGCTTCCTTTACCC 902  
Qy 274 AspAspGluGluLysGlnGlyLeuValAspGluLeuValLysLeuCysSerPheAspAsn 293  
Db 903 GAAGAAGAGGACCAAGAGGAGTTGTGAAGCTATCGCGAGCTGTGTAGCTTCGAGAAT 962  
Qy 294 LeuLysAsnGlnValAsnLysAsnGlySerSerTyrAsnSerLysLysLeuAspAsnLys 313  
Db 963 CTGAAGAAGTTGAGGAGTGAACAGTGAACAGTCGATCAAGAAC---TTTGAGATCGA 1019  
Qy 314 HisPhePheArgLysGlyGluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIle 333  
Db 1020 TTCTTGTTCGAAGGAGAGTGAAGTGAATGGGTAACTATTGTTCACCTTCAAGTG 1079  
Qy 334 LysLysLeu 336  
Db 1080 GAAAGATTG 1088  
RESULT 13  
AAC37672  
ID AAC37672 standard; DNA; 1154 BP.

XX AAC37672;  
XX AC 17-OCT-2000 (first entry)  
XX DT  
XX DE Arabidopsis thaliana DNA fragment SEQ ID NO: 18246.  
XX XX Hybridisation assay; genetic mapping; gene expression control;  
KW protein identification; signal transduction pathway; metabolic pathway;  
KW promoter; termination sequence; ss.  
XX Arabidopsis thaliana.  
XX OS  
XX PN EP1033405-A2.  
XX PD 06-SEP-2000.  
XX PF 25-FEB-2000; 2000EP-00301439.  
XX 25-FEB-1999; 99US-0121825P.  
PR 05-MAR-1999; 99US-0123180P.  
PR 09-MAR-1999; 99US-0123548P.  
PR 23-MAR-1999; 99US-0125788P.  
PR 25-MAR-1999; 99US-0126264P.  
PR 29-MAR-1999; 99US-0126785P.  
PR 01-APR-1999; 99US-0127462P.  
PR 06-APR-1999; 99US-0128234P.  
PR 08-APR-1999; 99US-0128714P.  
PR 16-APR-1999; 99US-0129845P.  
PR 19-APR-1999; 99US-0130077P.  
PR 21-APR-1999; 99US-0130449P.  
PR 23-APR-1999; 99US-0130510P.  
PR 23-APR-1999; 99US-0130891P.  
PR 28-APR-1999; 99US-0131449P.  
PR 30-APR-1999; 99US-0132048P.  
PR 30-APR-1999; 99US-0132407P.  
PR 04-MAY-1999; 99US-0132484P.  
PR 05-MAY-1999; 99US-0132485P.  
PR 06-MAY-1999; 99US-0132486P.  
PR 06-MAY-1999; 99US-0132487P.  
PR 07-MAY-1999; 99US-0132863P.  
PR 11-MAY-1999; 99US-0134256P.  
PR 14-MAY-1999; 99US-0134218P.  
PR 14-MAY-1999; 99US-0134219P.  
PR 14-MAY-1999; 99US-0134221P.  
PR 14-MAY-1999; 99US-0134370P.  
PR 18-MAY-1999; 99US-0134768P.  
PR 19-MAY-1999; 99US-0134941P.  
PR 20-MAY-1999; 99US-0135124P.  
PR 21-MAY-1999; 99US-0135353P.  
PR 24-MAY-1999; 99US-0135629P.  
PR 25-MAY-1999; 99US-0136021P.  
PR 27-MAY-1999; 99US-0136392P.  
PR 28-MAY-1999; 99US-0136782P.  
PR 01-JUN-1999; 99US-0137222P.  
PR 03-JUN-1999; 99US-0137528P.  
PR 04-JUN-1999; 99US-0137502P.  
PR 07-JUN-1999; 99US-0137724P.  
PR 08-JUN-1999; 99US-0138094P.  
PR 10-JUN-1999; 99US-0138540P.  
PR 10-JUN-1999; 99US-0138847P.  
PR 14-JUN-1999; 99US-0139119P.  
PR 16-JUN-1999; 99US-0139452P.  
PR 16-JUN-1999; 99US-0139453P.  
PR 17-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139454P.  
PR 18-JUN-1999; 99US-0139455P.  
PR 18-JUN-1999; 99US-0139456P.  
PR 18-JUN-1999; 99US-0139457P.  
PR 18-JUN-1999; 99US-0139458P.  
PR 18-JUN-1999; 99US-0139459P.  
PR 18-JUN-1999; 99US-0139460P.  
PR 18-JUN-1999; 99US-0139461P.



```
Db 174 ACTTCTCTACCTTACACAAAGATTTCATGGGTTATGGTCTCTACAACTACAAAGGTTGT 233
Qy 68 TTPGlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPhelysAlaArg 87
Db 234 TGGTACTATCCAAACACACTCCAAAGCGTCTTTCAGCGTCCAAACACACTTCAGGCCAGA 293
Qy 88 GluThrAspIleLeuThrLeuProLysAlaGlyThrThrThrThrThrThrThrThrThr 107
Db 294 GATCTGATATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 353
Qy 108 ThrPheAlaIleLeuThrArgAspValAsnHisProSerSerProThrHisProLeuLeu 127
Db 354 ATTTCCTCTGTGTATAGAGAAAGTACCGCGGACCCCTCAACACATCTTTGTCT 413
Qy 128 PhePheAsnProHisSerCysValGlnAsnLeuGluTyr---LeuTyrMetGlyArgGlu 146
Db 414 TTACAAACCCCTCATGACCTTGTCCCATTTCTTGTAGGTTGAGTTATAGCTTAATAGCCAA 473
Qy 147 AsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPheAlaGlyHisIle 166
Db 474 -----ATTCCGATCTCGCAAGTATCT---TCTCTATGATCTTTTCTACACATG 524
Qy 167 ProTyrSerLeuProAlaSerValLeuLysSerGlyThrLysIleLeuAsnIleSer 186
Db 525 CACTTACAAGCATTTGGTGAAGCACCACAAAGCT---TGCAAAACCCGTATATGTGT 581
Qy 187 ArgAsnArgLysSerThrPheValSerPheTyrLysPheGlyAsnLeuLeuAsnProAsp 206
Db 582 AGAGGTATCAAGATACGTTTCTCCGCTGTCATATAGAAACATGTTGATCGTCCACC 641
Qy 207 LysLeu-----LeuAspLeuGluLysSerValAspIlePheAlaSerGlyIleSerPhe 224
Db 642 AGATGGATCAAGCCACTTTGAGCTCATGTTTGTATGCTTATTGTAGAGGAGTCTCTTA 701
Qy 225 CysGlyProGluTrpAsnPheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSer 243
Db 702 TATGGACCTTATTTGGGAACATGTAATTGAGCTATTGGAAAGGGAGCTTGGAAACAAAGGAG 761
Qy 244 AsnLeuLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnValLysLys 263
Db 762 AATGTTCTTTTCATGAGTACGAGAGATTAATTGAGGAGCCTCGTTCCTCAAGTCAAGAGA 821
Qy 264 LeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGlnGlyIleValAsp 283
Db 822 CTCGCCGAGTCTTGGAAATGCCATTCACCAGGAAGAAAGAAAGTGGATCGGTGGAG 881
Qy 284 GluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnGlnValAsnLysAsnGly 303
Db 882 GAGATCTTGAAGTTGTAGTTTACGAATTTTACGAATTTTACGAATTTTACGAATTTTACGA 941
Qy 304 SerSerTyrAsnSerLysIleAspAsnLysHisPhePheArgLysGlyGluValArgAsp 323
Db 942 ACAAG---AGAATTGGTGTAGATTCTCAGGTCTCTTTAGGAAGAGTGAAGTTGTGTAT 998
Qy 324 TrpAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGlu 337
Db 999 TGAAGAATCATCTTACGCCACAAATGGCGCAACCTTTGAT 1040
RESULT 14
ID ABZ13222
XX ABZ13222 standard; DNA; 996 BP.
AC ABZ13222;
XX
XX
XX 21-JAN-2003 (first entry)
XX
XX Arabidopsis thaliana stress regulated gene SEQ ID NO 1027.
XX
XX Arabidopsis thaliana; plant; gene; stress; transgenic; ds.
XX
XX Arabidopsis thaliana.
XX
```

```
PN W0200216555-A2.
XX 28-FEB-2002.
XX
XX 24-AUG-2001; 2001WO-US026685.
XX
XX 24-AUG-2000; 2000US-0227866P.
XX
XX 26-JAN-2001; 2001US-0264647P.
XX
XX 22-JUN-2001; 2001US-0300111P.
XX
XX (SCRI ) SCRIPPS RES INST.
XX (SYGN ) SYNGENTA PARTICIPATIONS AG.
XX
XX Harper JF, Krepes J, Wang X, Zhu T;
XX
XX WPI; 2002-304127/34.
XX
XX Identifying a stress condition to which a plant cell has been exposed and
XX producing plants with increased tolerance to these abiotic stresses.
XX
XX Claim 144; SEQ ID NO 1027; 577bp + Sequence Listing; English.
XX
XX The invention relates to identifying a stress condition to which a plant
XX cell has been exposed, comprising: (a) contacting nucleic acid
XX representative of expressed polynucleotides in the plant cell with an
XX array or probes representative of the plant cell genome; and (b)
XX detecting a profile of expressed polynucleotides in the plant cell
XX characteristic of a stress response. The method is useful in the
XX production of transgenic plants, cells and seeds and in producing plants
XX with increased tolerance to abiotic stress. The present sequence is that
XX of an Arabidopsis thaliana stress regulated gene (ABZ12196-ABZ17574) used
XX in methods of the invention. Note: The sequence data for this patent is
XX not represented in the printed specification but is based on sequence
XX information supplied to Derwent by the European Patent Office
XX
XX SQ Sequence 996 BP; 306 A; 199 C; 213 G; 278 T; 0 U; 0 Other;
XX
XX Alignment Scores:
XX Pred. No.: 1,34e-43 Length: 996
XX Score: 545.00 Matches: 118
XX Percent Similarity: 58.31% Conservative: 61
XX Best Local Similarity: 38.44% Mismatches: 118
XX Query Match: 26.47% Indels: 10
XX DB: 6 Gaps: 8
XX
XX US-09-854-122-16 (1-391) x ABZ13222 (1-996)
XX
XX Qy 36 GluAspSerLysMetTyrLysArgTyrArgLysValSerSerLeuProSerAsn--- 54
XX
XX Db 43 AAAGCGACACACGTTAGTCAAGAAACCAAGAACTTGATCACTTCTCTACCTTCAGACAAA 102
XX
XX Qy 55 AspTyrTrpGlyAspThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuVal 74
XX
XX Db 103 GATTTCATGGGTTATGGTCTCTCAACACTACAAAGGTTGGTGGTACTATCCAAACACACTC 162
XX
XX Qy 75 ProGlyIleMetAlaPheGluAspAsnPhelysAlaArgGluThrAspIleLeuThr 94
XX
XX Db 163 CAAGCGGTCTTGAGTCCAAACACACTTCAGCCACGAGATCTGATATAATCCTCGCT 222
XX
XX Qy 95 ThrLeuProLysAlaGlyThrThrThrThrThrThrThrThrThrThrThrThrThrThr 114
XX
XX Db 223 TCTTTGCCAAAGGTGGAACCACTTGGCTCAAAATCCCTTAATTTTCGTGTGTACATAGA 282
XX
XX Qy 115 AspValAsnHisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCys 134
XX
XX Db 283 GAAAGTACCCGGGAACCCCTCAACACATCTTTGTCTTACAAAACCCCTCATGACCTT 342
XX
XX Qy 135 ValGlnAsnLeuGluTyr---LeuTyrMetGlyArgGluAsnThrMetProAspLeuAsp 153
XX
XX Db 343 GTCCCATTTCTTGAGTTGAGTTATAGCTTAATAGCCAA-----ATTCCGGATCTCGCA 396
XX
XX Qy 154 MetLeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAla 173
XX
XX
```



Db	736	...             : : : : :	-----GTGAGAGAGATTGTGAAG	783
QY	288	LeuCyssSerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyrAsn	307	
Db	784	TTGTGTAGCTTTGAGAGTTTAAGTAATTTGCGAAGTTAACAAAGAGGGAAATTCGCAAA	843	
QY	308	SerLysIleAspAsnLysHisPhePheArgLysGlyValArgAspTrpAlaSerTyr	327	
Db	844	GGA---ATAGAGACTAATAACTTTCTTGAAGAAGGAGAGATTGGAGGATGGAGAGAT	900	
QY	328	LeuThrSerGluMetIleLysLysLeuGlu	337	
Db	901	TTGAGTGAGTCAITGGCAGAGAGAAATTCAT	930	

Search completed: March 20, 2004, 04:00:57  
Job time : 518 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 20, 2004, 03:46:04 ; Search time 97 seconds  
(without alignments)  
2236.967 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059  
Sequence: 1 TRGTGTAACVACFLIMAGIL.....CLLSYFVNNFILLKKKK 391

Scoring table:

BLOSUM62  
Xgapop 10.0, Xgapext 0.5  
Ygapop 10.0, Ygapext 0.5  
Fgapop 6.0, Fgapext 7.0  
Delop 6.0, Delext 7.0

Searched: 682709 segs, 277475446 residues

Total number of hits satisfying chosen parameters: 1365418

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODEL=frame\_p2n.model -DEV=xlh  
-Q/cgn2\_1/USPTO.spool/US09854122/runat\_18032004\_100001\_3276/app\_query.fasta\_1.583  
-DB=Issued Patents NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=ptc -NCORE=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-USER=US09854122@cgn\_1\_1\_56@runat\_18032004\_100001\_3276 -NCPUS=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

Issued Patents NA:\*  
1: /cgn2\_6/prodata/2/ina/5A COMB.seq:\*  
2: /cgn2\_6/prodata/2/ina/5B COMB.seq:\*  
3: /cgn2\_6/prodata/2/ina/5A COMB.seq:\*  
4: /cgn2\_6/prodata/2/ina/5B COMB.seq:\*  
5: /cgn2\_6/prodata/2/ina/5A COMB.seq:\*  
6: /cgn2\_6/prodata/2/ina/5B COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	273	13.3	1063	1	US-08-325-562-1
2	273	13.3	1063	1	US-08-437-795-1
3	268	13.0	1004	4	US-09-609-816-3
4	264	12.8	942	4	US-09-609-816-1
5	260	12.6	922	4	US-09-609-816-2
6	239	11.6	798	4	US-09-795-926-5
7	239	11.6	912	4	US-09-795-926-1
8	228	11.1	1366	4	US-09-328-1748-2
9	209.5	10.2	2153	4	US-09-795-926-19
10	200	9.7	1564	2	US-08-852-481-1
11	187	9.1	1130	4	US-09-786-240-22
12	184	8.9	645	4	US-09-795-926-15

13	155.5	7.6	447	4	US-09-795-926-7	Sequence 7, Appli
14	155.5	7.6	561	4	US-09-795-926-11	Sequence 11, Appli
15	124	6.0	3309	4	US-09-510-543-20	Sequence 20, Appli
16	124	6.0	3614	4	US-09-023-655-968	Sequence 968, App
17	124	6.0	5878	4	US-09-510-543-18	Sequence 18, Appli
18	119	5.8	5181	1	US-08-357-073-10	Sequence 10, Appli
19	117	5.7	3648	1	US-08-053-614-1	Sequence 1, Appli
20	117	5.7	3648	1	US-08-316-397B-1	Sequence 1, Appli
21	117	5.7	3648	2	US-09-034-306-1	Sequence 1, Appli
22	117	5.7	3648	3	US-09-259-437-1	Sequence 1, Appli
23	117	5.7	3648	5	PCT-US93-09782-1	Sequence 3, Appli
24	117	5.7	4821	1	US-08-053-614-3	Sequence 3, Appli
25	117	5.7	4821	1	US-08-316-397B-3	Sequence 3, Appli
26	117	5.7	4821	2	US-09-034-306-3	Sequence 3, Appli
27	117	5.7	4821	3	US-09-259-437-3	Sequence 3, Appli
28	117	5.7	4821	5	PCT-US93-09782-3	Sequence 3, Appli
29	111	5.4	333	4	US-09-795-926-3	Sequence 3, Appli
30	111	5.4	447	4	US-09-795-926-9	Sequence 9, Appli
31	111	5.4	5925	3	US-08-470-260-4	Sequence 4, Appli
32	111	5.4	5925	3	US-08-471-491-4	Sequence 4, Appli
33	111	5.4	5925	3	US-08-466-662-4	Sequence 4, Appli
34	111	5.4	10299	2	US-08-477-451-1	Sequence 1, Appli
35	111	5.4	10299	2	US-08-477-451-5	Sequence 5, Appli
36	111	5.4	19932	2	US-08-477-451-25	Sequence 25, Appli
37	111	5.4	640681	4	US-09-790-988-1	Sequence 1, Appli
38	105	5.1	129908	4	US-09-585-858-1	Sequence 1, Appli
39	104	5.1	41708	4	US-09-470-512A-3	Sequence 3, Appli
40	103	5.0	4404	4	US-09-134-000C-3335	Sequence 3335, Ap
41	103	5.0	11015	4	US-10-304-708-55	Sequence 55, Appli
42	102	5.0	580073	4	US-08-545-528D-1	Sequence 1, Appli
43	100.5	4.9	294	4	US-09-795-926-17	Sequence 17, Appli
44	99.5	4.8	3385	4	US-09-509-800-1	Sequence 1, Appli
45	99.5	4.8	9636	1	US-08-323-170B-1	Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-08-325-562-1  
; Sequence 1, Application US/08325562  
; Patent No. 5714594  
; GENERAL INFORMATION:  
; APPLICANT: Weinshilboum, Richard M.  
; APPLICANT: Aksoy, Ibrahim A.  
; APPLICANT: Wood, Thomas C.  
; TITLE OF INVENTION: CDNA CLONING AND EXPRESSION OF HUMAN  
; TITLE OF INVENTION: LIVER ESTROGEN SULFOTRANSFERASE  
; NUMBER OF SEQUENCES: 17  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Schwegman, Lundberg & Woessner P.A.  
; STREET: 3500 IDS Center  
; CITY: Minneapolis  
; STATE: MN  
; COUNTRY: USA  
; ZIP: 55402  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/325,562  
; FILING DATE: 18-OCT-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Mueeting, Ann M.  
; REGISTRATION NUMBER: 33,977  
; REFERENCE/DOCKET NUMBER: 150.131US1  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 612-339-0331  
; TELEFAX: 612-339-3061  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:

```

; LENGTH: 1063 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 107..989
US-08-325-562-1

Alignment Scores:
Pred. No.: 7,06e-23 Length: 1063
Score: 273.00 Matches: 99
Percent Similarity: 42.62% Conserv: 57
Best Local Similarity: 27.05% Mismatches: 140
Query Match: 13.26% Indels: 70
DB: 1 Gaps: 16

US-09-854-122-16 (1-391) x US-08-325-562-1 (1-1063)

QY 41 TyrLysArgTyrArgGluLeuValSerSerLeuProSerAsnAspTyrTrpGlyAspThr 60
DB 128 TATGAAAGTTTGAGAGAGTCCATGGGATCTA-----160
QY 61 MetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyLeuMetAlaPhe 80
DB 161 -----ATGTATAAGATTT-----GTCAATAT 184
QY 81 GluAspAsn-----PheValaArgGluThrAspIleLeuThrLeuPro 97
DB 185 TGGGATATGTGGAGCGTTTCAGCAGACACCATGATCTGTGTCATGTCACCTACCT 244
QY 98 LysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaLeuThrArg-----AspVal 116
DB 245 AAATCTGTGACAACTGGCTGTAGTCAAAATCTGTATATGATCTATAAGAGGGTGTGTG 304
QY 117 AsnHisProSerProThrHisProLeuLeuPheAsnProHisSerCysValGln 136
DB 305 GAAAGTGCAGAAAGAT-----GTAATTTTAAAT-----334
QY 137 AsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn 156
DB 335 CGAATACCTTCTCGAATGCAGAAAGAAACCTCATGATGGAGTAAACAAATAGAT 394
QY 157 Glu-----SerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSer 174
DB 395 GAGATGAATCTCTAGAACTGTGAAGACTCATTTGCCACCTGAACTCTTCTCCCTCCCA 454
QY 175 ValLeuLysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheVal 194
DB 455 TTTTGGGAAAGAGATTGTAAAGATATCTATCTTCCCGGAATGCCAAGGATGTGCTGT 514
QY 195 SerPheTrpLysPheGlyAsnLeuLe-----AsnProAspLysLeuLeuAsp 210
DB 515 TCCTTTTATTATTTCTTTCTAATGGTGGTGTCTATCTCCCAATCTCGGATCTTTCCAGAG 574
QY 211 LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsn 230
DB 575 TTT-----GTGAGAAATTCATGACGACGAGGTCTCTATGTT-----TCCTGGAT 622
QY 231 PheGlnAlaGluPheThrAsnAlaLaSerThrAsnSerAsnLeuLeuLeuSerTyr 250
DB 623 AAACATGTAATAATCTTGTGGGAAAGGAAAGATCCACGTGACTATTTCTTTCTAC 682
QY 251 GluGluMetLeuGluLysProValGluAsnValLysLeuAlaGluPheMetGlyCys 270
DB 683 GAAGACCTGAAGAGATATTCAGAAAGAGGTGATAAATTTGATACATTTCTCTG-----736
QY 271 GlyPheThrAspAspGluLysGlnGlyIleValAspGluIleValLysLeuCysSer 290
DB 737 -----GAAAGGAAGCCATCAGAGGAGCTGTGACAGGATTTATACATCATCTCTCG 787
QY 291 PheAspAsnLeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIle 310

```

```

DB 788 TTCCAAGAGATGAAGAAATCCCA---TCCCAAAATTACACAACTGCCAGCAAAATT 844
QY 311 AsnAsnLysHis-----PhePheArgGlyGluValArgAspTrpAlaAsnTyr 327
DB 845 ATGAACCAAGAAATGTGCCCTTTCATGAGAAGGGAATTACAGAGACTGGAAAAATCAC 904
QY 328 LeuThrSerGluMetIleLysLysLeuGluThrAlaGlyLysIleAsnGluSerGluLys 347
DB 905 TTTACAGTAGCCCTGAATGAAAAATTT-----GATAAA 937
QY 348 HisLeuLeuSerAsnLysAsnLeuThrAsnPheAsnLeuAsnTyrCysGluLysSer 367
DB 938 CATTTAGACGACCAATGAAGAAATCTACACTGAAGTTT-----CGAAT 982
QY 368 AsnIle-----SerLeuCysLeuLeuSerTyrSerPheValIleAsnAsnPheIleLeu 385
DB 983 GAGATCTAAGAGGTCTTTCTTAAACATATCTGATATTAAGATTTCTTTCTTCA 1042
QY 386 LeuLysLysLysLysLysLys 391
DB 1043 TTTCAAAAAAATAAAAAA 1060

RESULT 2
US-08-437-795-1
; Sequence 1, Application US/08437795
; Patent No. 5744355
; GENERAL INFORMATION:
; APPLICANT: Weinshilbom, Richard M.
; APPLICANT: Aksoy, Ibrahim A.
; APPLICANT: Wood, Thomas C.
; TITLE OF INVENTION: CDNA CLONING AND EXPRESSION OF HUMAN
; TITLE OF INVENTION: LIVER ESTROGEN SULFOTRANSFERASE
; NUMBER OF SEQUENCES: 19
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Schwegman, Lundberg & Woessner P.A.
; STREET: 3500 IDS Center
; CITY: Minneapolis
; STATE: MN
; COUNTRY: USA
; ZIP: 55402
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/437,795
; FILING DATE: 09-MAY-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Mueeting, Ann M.
; REGISTRATION NUMBER: 33,977
; REFERENCE/DOCKET NUMBER: 150.137US1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 612-339-0331
; TELEFAX: 612-339-3061
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1063 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 107..989
US-08-437-795-1

Alignment Scores:
Pred. No.: 7,06e-23 Length: 1063
Score: 273.00 Matches: 99
Percent Similarity: 42.62% Conserv: 57

```

```

Best Local Similarity: 27.05%      Mismatches: 140
Query Match:          13.26%      Indels:      70
DB:                   1          Gaps:        16

US-09-854-122-16 (1-391) x US-08-437-795-1 (1-1063)

QY 41 TYRLeuArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTrpGlyAspThr 60
Db 128 TATGAAGATTGGAAGAGTCATGGATTCTA-----160
QY 61 MetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAlaPhe 80
Db 161 -----ATGTATATAAGATTT-----GTCAATAT 184
QY 81 GluAspAsn-----PheLysAlaArgGluThrAspIleLeuThrLeuPro 97
Db 185 TGGGATATGTGAAGCGTTCAGGACAGACAGATGATGTGTCATTCGCCACCTACCT 244
QY 98 LysAlaGlyThrTrpThrLysAlaLeuThrPheAlaLeuThrArg---AspVal 116
Db 245 AAATCTGGTACAACTGGGTAGTGAATTTGTATATGATCTATAAAGAGGGTATGTG 304
QY 117 AsnHisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGln 136
Db 305 GAAAGTGCARAAGAT-----GTAATTTTAAT-----334
QY 137 AsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn 156
Db 335 CGAATACCTTCTCGAATGCAGAAAGAAACCTCATGAATGGAGTAAACAAATAGAT 394
QY 157 Glu-----SerProArgLeuPheAlaGlyHisIleProTyrSerLeuProAlaSer 174
Db 395 GAGATGAATTCCTAGAATTTGAAGACTCAITGGCCACCTGAACCTTCCTCCCTCA 454
QY 175 ValLeuLysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheVal 194
Db 455 TTTTGGGAAAAGGATTGTAAGATACTATCTTTGCCGGAATGCAAGGATGGCTGT 514
QY 195 SerPheTrpLysPheGlyAsnLeuIle-----AsnProAspLysLeuLeuAsp 210
Db 515 TCTTTTATATTCTTCTTAATGGTGGTGGTGCATCCAAATCCTGGATCCTTTCCAGAG 574
QY 211 LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsn 230
Db 575 TTT-----GTGGAGAAATTCATGCAAGACAGAGTTCCCTTATGTT--TCCTGGTAT 622
QY 231 PheGluAlaGluPheThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyr 250
Db 623 AAACATGTAAATCTTGGTGGGAAAAGGGAAGAGTCCACGTGACTATTTCTTTCTAC 682
QY 251 GluGluMetLeuGluLysProValGluAsnValLysLeuAlaGluPheMetGlyCys 270
Db 683 GAAGACCTGAAGAGGATATCAGAAAGAGGTGATAAATTTGATACATTCCTG-----736
QY 271 GlyPheThrAspAspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSer 290
Db 737 -----GAAAGGAAGCCATCAGAGAGCTTGTGGACAGGATTATACATCATCTCG 787
QY 291 PheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIle 310
Db 788 TTCAGAGATGAAGAAATCA-----TCCAAATTTACACACATGCCAGCAGAAAT 844
QY 311 AspAsnLysHis-----PhePheArgLysGlyGluValArgAspTrpAlaAsnTyr 327
Db 845 ATGAACCCAGAAATTTGCGCCCTTCATGAGAAGGAATACAGGAGCTGGAATAATCAC 904
QY 328 LeuThrSerGluMetIleLysLeuGluThrAlaGlyLysIleAsnGluSerGluLys 347
Db 905 TTTACAGTACCCCTGAATGAAATTT-----GATAAA 937
QY 348 HisLeuLeuSerAsnLysAsnLeuThrAsnPheAsnLeuAsnTyrCysGluLysSer 367
Db 938 CATTTATGAGCAGCAATGAAGGAATCTACACTGAAGTTT-----CGAAGT 982

368 AsnIle-----SerLeuCysLeuLeuSerTyrSerPheValIleAsnAsnPheIleLeu 385
Db 983 GAGATCTAAGAAGTCTTTCTTTACTTTACATATCTGATATTAAAGATTTCTTTTCATTA 1042
QY 386 LeuLysLysLysLysLys 391
Db 1043 TTTCAAAAAAATAAAAAA 1060

RESULT 3
US-09-609-816-3
; Sequence 3, Application US/09609816
; Patent No. 643684
; GENERAL INFORMATION:
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: DiFrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 1004
; TYPE: DNA
; ORGANISM: HUMAN
US-09-609-816-3

Alignment Scores:
Pred. No.: 2,568-22      Length: 1004
Score: 268.00      Matches: 97
Percent Similarity: 43.97%      Conservative: 56
Best Local Similarity: 27.87%      Mismatches: 131
Query Match: 13.02%      Indels: 64
DB: 4      Gaps: 18

US-09-854-122-16 (1-391) x US-09-609-816-3 (1-1004)
QY 29 LysAsnGluGlnGluLysGluGluAspSerLysMetTyrLysArgTyrArgGluIleVal 48
Db 62 AAAAAACGCTCCACGATCGAAAAAAGCCAGAACTGTTTAAACATCATCGAA-----GTA 115
QY 49 SerSerLeuProSerAsnAspTyrTrpGlyAspThrMetArgLeuTyrLysGlyPheTrp 68
Db 116 GATGGAGTCCCT-----ACGTGTATATTATCAAAAGATGTGG 154
QY 69 GlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPheLysAlaArgGlu 88
Db 155 GAAAAA-----GTCTGTAATTTCCCAAGCCAGCCT 184
QY 89 ThrAspIleLeuThrThrLeuProLysAlaGlyThrThrTrpThrLysAlaLeuThr 108
Db 185 GATGATCTTATCTGGCAACTTACCCAAAGTCAGGTACACATGGATGCGATGAATTTTA 244
QY 109 PheAlaIleLeuThr---ArgAspValAsnHis-----ProSerSerProThr 123
Db 245 GACATGATCTTAATGATGGTGAATGGAGAAATGCAAAAGAGCCAGACTCTAGATAGA 304
QY 124 HisProLeuLeuPhePheAsn---ProHisSerCysValGlnAsnLeuGluTyrLeuTyr 142
Db 305 CACGCTTTCCTTGAACCTGAAATTTCCCAATAAAGAAAAACAGATTTGGAGTTTCGTT--- 361

```

```
143 MetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPhe 162
362 -----CTTGAATG-----TCCTCACCACACTGATA 388
163 AlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIle 182
389 AAAACACATCTCCTTCCATCTGATTCACCACTCTCTCGGAAGAAACATGCAAGATT 448
183 IleAsnIleSerArgAsnAlaGlySerThrPheValSerPheTyrLysPheGlyAsnLeu 202
449 GTCTATGTGCCAGGAATCCAGGATGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCT 508
203 IleAsn-----ProAspLysLeuLeuAspLeuLysSerValAspIlePheAlaSer 220
509 GCTTCCTTTATGCTGATCCT---CAGAACTTACAGGAATTTTATGAGAAATTCATGTC 565
221 GlyIleSerPheCysGlyProGluTyrAsnPheGluAlaGluPheThrAsnAlaAlaSer 240
566 GGA---AAAGTTGTGGGGTCTGCTGTTGACCATGTGAAGATGCTGGGCTGCAGAA 622
241 ThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuLysProValGluAsn 260
623 GACACGACCGGATCCTCTACCTCTTCTACGAGGATATTAATAAATAATCCAAACATGAG 682
261 ValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGlnGly 280
683 ATCCACAGGTTGGAAATCTTGGAGAAACCTTTGTCAGTGTAT-----727
281 IleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnValAsn 300
728 GTTATAAACAAAGATTGTCACCATACCTCTTGTATGTAATGAAGGATATCCCATG---784
301 LysAsnGlySerSerTyrAsnSerLysIleAspAsn-----LysHisPhePheArg 317
785 GCCAACCATATCGGTACCTGCTCACAATATTCATCTCATCTCCATCTCAAAATTATGAGG 844
318 LysGlyGluValAlaGAspTyrAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGlu 337
845 AAAGGGATGCTGGAGACTGGAAGAACCATCTTACTGTGGCTATG-----889
338 ThrAlaGlyLysIleAsnGluSer---GluLysHisLeuLeuSerAsnLysAsnLeuThr 356
890 -----AATGAGAACTTTGATAAGCATAT-----GAAGAAGATGGCA 928
357 AsnPheAsnLeuAsnAsnTyrCys 364
929 GGGTCCACATG---AATCTCTGC 949
```

```
RESULT 4
US-09-609-816-1
; Sequence 1, Application US/09609816
; Patent No. 6436694
; GENERAL INFORMATION:
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: DiFrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 1
```

```
LENGTH: 942
TYPE: DNA
ORGANISM: HUMAN
US-09-609-816-1
Alignment Scores:
Pred. No.: 6,97e-22
Score: 264.00
Percent Similarity: 44.25%
Best Local Similarity: 27.59%
Query Match: 12.82%
Gaps: 4
DB: 18

US-09-854-122-16 (1-391) x US-09-609-816-1 (1-942)
QY 29 LysAsnGluGlnGluLysGluLysSerLysMetTyrLysArgTyrArgGluIleVal 48
Db 23 AAAACCGCTCCACGATGGAAGAAAGCCAGACTGTTTAAACATCATGGA---GTA 76
QY 49 SerSerLeuProSerAsnAspTyrTyrGlyAspThrMetArgLeuTyrLysGlyPheTyr 68
Db 77 GATGGAGTCCCT-----ACGTTGATATTATCAAAAGAAATGSGTG 115
QY 69 GlnMetGlyTyrLeuValProGlyIleValPheGluAspAsnPheLysAlaArgGlu 88
Db 116 GAAAA-----GTATGTAATTTCCAAAGCCAGCCT 145
QY 89 ThrAspIleIleLeuThrThrLeuProLysAlaGlyThrThrTyrThrLysAlaLeuThr 108
Db 146 GATGATCTTATTTCTGGCACTTACCCAAAGTCAGGTACCAACATGATCATGAAATTTTA 205
QY 109 PheAlaIleLeuThr---ArgAspValAsnHis-----ProSerSerProThr 123
Db 206 GACATGATTTCTAAATGATGTGTGGAGAAATTCAAAAAGAGCCAGACTCTAGATAGA 265
QY 124 HisProLeuLeuPhePheAsn---ProHisSerCysValGlnAsnLeuGluTyrLeuTyr 142
Db 266 CACGCTTCTCTGAACTGAAATTTCCCATTAAGAAAAACCCAGATTTCGATTCGTT---322
QY 143 MetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPhe 162
Db 323 -----CTTGAATG-----TCCTCACCACACTGATA 349
QY 163 AlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIle 182
Db 350 AAAACACATCTCCTTTCATCTGATTCACCATCTATCTGGAAGAAACATGCAAGATT 409
QY 183 IleAsnIleSerArgAsnArgLysSerThrPheValSerPheTyrLysPheGlyAsnLeu 202
Db 410 GTCTATGTGCCAGAAATCCCAAGGATTCCTGCTGCTCTACTACCATCTTTCACAGGATG 469
QY 203 IleAsn-----ProAspLysLeuLeuAspLeuLysSerValAspIlePheAlaSer 220
Db 470 GCTTCCTTTATGCTGATCCT---CAGAACTTAGAGGAATTTTATGAGAAATTCATGTC 526
QY 221 GlyIleSerPheCysGlyProGluTyrAsnPheGlnAlaGluPheThrAsnAlaAlaSer 240
Db 527 GGA---AAAGTTCTTGGCGGCTCTGTTGACCATGTGAAGGATGTGGGCTGCAAAA 583
QY 241 ThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuLeuLysProValGluAsn 260
Db 584 GACATGACCGGATCTCTCTACCTCTTCTACGAGGATATTAAAAAATCCAAACATGAG 643
QY 261 ValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluLysGlnGly 280
Db 644 ATCCACAGGTGTGGAATTTCTGGAGAAACCTTGTGTCAGGTGAT-----688
QY 281 IleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnValAsn 300
Db 689 GTTATAAACAAAGATTGTCACCATCTCTCTGATGTAATGAAGGATATTCATGTCATG 745
QY 301 LysAsnGlySerSerTyrAsnSerLysIleAspAsn-----LysHisPhePheArg 317
```

```
Db 746 GCCAACCATCTGCGGTACTGTCACATATTCAATCACTCCATCTCATAAATTTATGAGG 805
Qy 318 LysGluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLysLysLeuGlu 337
Db 806 AAAGGGATGCTGAGACTGGAGAAACCACTTTACTGTG-----844
Qy 338 ThrAlaGlyLysIleAsnGluSer---GluLysHisLeuLeuSerAsnLysAsnLeuThr 356
Db 845 -----GCTTGTGATGAGACTTTGATAGCAATTAT-----GAAAGAAGATGGCA 889
Qy 357 AsnPheAsnLeuAsnAsnTyrCys 364
Db 890 GGGTCCACACTG---AACTTCTGC 910

RESULT 5
US-09-609-816-2
; Sequence 2, Application US/09609816
; Patent No. 643684
; GENERAL INFORMATION:
; APPLICANT: Woodage, Trevor
; APPLICANT: Wei, Minh Hui
; APPLICANT: Kodira, Chinappa
; APPLICANT: Beasley, Ellen
; APPLICANT: Difrancesco, Valentina
; TITLE OF INVENTION: ISOLATED HUMAN DRUG-METABOLIZING
; TITLE OF INVENTION: PROTEINS, NUCLEIC ACID MOLECULES ENCODING HUMAN
; TITLE OF INVENTION: DRUG-METABOLIZING PROTEINS, AND USES THEREOF
; FILE REFERENCE: CL000669PCT
; CURRENT APPLICATION NUMBER: US/09/609,816
; CURRENT FILING DATE: 2000-07-03
; PRIOR APPLICATION NUMBER: 60/192,408
; PRIOR FILING DATE: 2000-03-27
; PRIOR APPLICATION NUMBER: 60/212,725
; PRIOR FILING DATE: 2000-06-20
; PRIOR APPLICATION NUMBER: 09/609,816
; PRIOR FILING DATE: 2000-07-03
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 922
; TYPE: DNA
; ORGANISM: HUMAN
US-09-609-816-2

Alignment Scores:
Pred. No.: 2,03e-21 Length: 922
Score: 260.00 Matches: 86
Percent Similarity: 45.02% Conservative: 54
Best Local Similarity: 27.65% Mismatches: 123
Query Match: 12.63% Indels: 48
DB: 4 Gaps: 13

US-09-854-122-16 (1-391) x US-09-609-816-2 (1-922)
Qy 29 LysAsnGluGlnGluLysGluAspSerLysMetTyrLysArgTyrArgGluIleVal 48
Db 23 AAAAAAGCTCCACGATGAGAAAGAAAGCCAGAACTGTTTAAATCATCATGGAA-----GTA 76
Qy 49 SerSerLeuProSerAsnAspTyrTyrGlyAspThrMetArgLeuTyrLysGlyPheTyr 68
Db 77 GATGAGTGCCT-----ACGTTGATATATATCAAAAGAAATGGTGG 115
Qy 69 GlnMetGlyTyrLeuValProGlyIleMetAlaPheGluAspAsnPhelYsAlaArgGlu 88
Db 116 GAAAA-----GTATGTAATTTCCAAGCAAGCCT 145
Qy 89 ThrAspIleLeuThrThrLeuProLysAlaGlyThrThrThrThrLysAlaLeuThr 108
Db 146 GATGATCTTATCTGGCAACTTACCCAAAGTCAGGTACAAACATGATGATGAATTTTA 205
Qy 109 PheAlaIleLeuThr---ArgAspValAsnHis-----ProSerProThr 123
Db 206 GACATGATTTAAATGATGGTGTGAGAAATGCGAAAGAGCCAGACTCTAGATAGA 265
```

```
Qy 124 HisProLeuLeuPhePheAsn---ProHisSerCysValGlnAsnLeuGluTyrLeuTyr 142
Db 266 CACCTTCTTCTGAACTGAAATTTCCCATTAAGAAAAACACAGATTGGAGTTGTT---322
Qy 143 MetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPhe 162
Db 323 -----CTTCAATG-----TCCTCACCACCACTGATA 349
Qy 163 AlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThrLysIle 182
Db 350 AAAACACATCTCCCTTCATCTGATCTCCACCATCTATCTGGAAGAAAACTGCAAGATT 409
Qy 183 IleAsnIleSerArgAsnArgLysSerThrPheValSerPheThrLysPheGlyAsnLeu 202
Db 410 GTCTATGTGGCCAGAAATCCACAGGATTGCTGTGTCTACTACTACCACTTTACAGGATG 469
Qy 203 IleAsn-----ProAspLysLeuLeuAspLeuLysSerValAspIlePheAlaSer 220
Db 470 GCTTCTCTTTATGCTGCTGCTCT---CAGAACTTAGAGGAATTTTATGAGAAATTCATGTC 526
Qy 221 GlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPheThrAsnAlaAlaSer 240
Db 527 GGA---AAAGTTGTTGGGGCTCCTGTTGACCATGTGAAAGGATGCTGGCTGCAAAA 583
Qy 241 ThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsn 260
Db 584 GACATGACACCGGATCTCTACCTCTTCTACGAGGATATTAATAAGAACCCAAACGGGAA 643
Qy 261 ValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluLysGlnGly 280
Db 644 ATTGAGAAGATACCTGAAGTTCTGTG-----GAAAAAGACATATCAGAGGAA 688
Qy 281 IleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnValAsn 300
Db 689 ATTCTGATATAAATCATCATCATCACCTCTTCTGAGGATTAATGAAGCAAAACCAATGACC 748
Qy 301 LysAsnGlySerSerTyrAsnSerLysIleAspAsn-----LysHisPhePheArgLys 318
Db 749 AACTATACCATCTTGGCCACCCAGCATTTATGGACCATCTCCCTCTTTATGAGGAA 808
Qy 319 GlyGluValArgAspTrpAlaAsnTyrLeuThr 329
Db 809 GGGATCGCTGGAGACTGGAAGAACTATTTTACT 841

RESULT 6
US-09-795-926-5
; Sequence 5, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Donoho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
```







```

US-09-795-926-19
; Sequence 19, Application US/09795926
; Patent No. 6555669
; GENERAL INFORMATION:
; APPLICANT: Doncho, Gregory
; APPLICANT: Hilbun, Erin
; APPLICANT: Turner, C. Alexander Jr.
; APPLICANT: Friedrich, Glenn
; APPLICANT: Abuin, Alejandro
; APPLICANT: Zambrowicz, Brian
; APPLICANT: Sands, Arthur T.
; APPLICANT: Walke, D. Wade
; APPLICANT: Wilganowski, Nathaniel L.
; APPLICANT: Hu, Yi
; APPLICANT: Kieke, James Alvin
; APPLICANT: Potter, David George
; TITLE OF INVENTION: POLY HUMAN TRANSFERASE PROTEINS AND
; TITLE OF INVENTION: POLYNUCLEOTIDES ENCODING THE SAME
; FILE REFERENCE: LEX-0144-USA
; CURRENT APPLICATION NUMBER: US/09/795,926
; CURRENT FILING DATE: 2001-02-28
; PRIOR APPLICATION NUMBER: US 60/185,920
; PRIOR FILING DATE: 2000-02-29
; PRIOR APPLICATION NUMBER: US 60/186,558
; PRIOR FILING DATE: 2000-03-02
; PRIOR APPLICATION NUMBER: US 60/191,849
; PRIOR FILING DATE: 2000-03-24
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 19
; LENGTH: 2153
; TYPE: DNA
; ORGANISM: homo sapiens
US-09-795-926-19

Alignment Scores:
Pred. No.: 8,88e-15 Length: 2153
Score: 209.50 Matches: 82
Percent Similarity: 37.04% Conservative: 48
Best Local Similarity: 23.36% Mismatches: 123
Query Match: 10.17% Indels: 98
DB: 4 Gaps: 13

US-09-854-122-16 (1-391) x US-09-795-926-19 (1-2153)
QY 82 AspaasnPhelYsAlaAxsGluThrApIlelleLeuThrLeuProlysaGlyThr 101
Db 678 GACACCTTCGAAGCCACACATGATGACATCGCTAGCATCTATCCAAAGTGGGTTC 737
QY 102 ThrTrp-----ThrLysAlaLeuThrPheAlaIleLeuThrArgaspValasn 117
Db 738 AACTGGATTCACCATTTGTCAGTGAATTAATATATGCTGTTCTTCTAAAAAAGTATAA 797
QY 118 HisProSerSerPro----- 122
Db 798 TATCCAGAAATCCCAGTCTTGAATCTGGGATTCAGAAAATATCTAGTCACAGTAGTG 857
QY 123 -----ThrHisProLeuLeu 127
Db 858 CTAGAAACTATCACTTAGATATACCAAACTAGAGTAGTATCAACACACACACACA----- 911
QY 128 PhePheAsnProHisSerCyseValGlnAsnLeuGluTyLeuTyT-MetGlyArgGluAs 147
Db 912 -----CATGCACACACGTCGCACATTTTCACCTTTTGTGTATATTTTAAAGAGAAATGAA 965
QY 147 nThrMetProAspLeuaspMetLeuasnGluSerProArgLeuPheAlaGlyHisIlePr 167
Db 966 AGGCTTTTCCA-----TCACCAAGGATTTTGGCAACTTCACCTCCA 1004
QY 167 ofYrsSerLeuLeuProLaserValLeuIlysSerGlyThrLys----- 181
Db 1005 CTATACACAAATACTCGGCTTATCTTCGAGATAAAGCCAGACAGACAGCATCTCAT 1064

```







Db 604 CTCGGAGCAAGTTGAAGTATGATCATATTCGCCAG 639

RESULT 13

US-09-795-926-7

; Sequence 7, Application US/09795926

; Patent No. 6555669

; GENERAL INFORMATION:

; APPLICANT: Donoho, Gregory

; APPLICANT: Hilbun, Erin

; APPLICANT: Turner, C. Alexander Jr.

; APPLICANT: Friedrich, Glenn

; APPLICANT: Abuin, Alejandro

; APPLICANT: Sambrowicz, Brian

; APPLICANT: Sands, Arthur T.

; APPLICANT: Walke, D. Wade

; APPLICANT: Wilganowski, Nathaniel L.

; APPLICANT: Hu, Yi

; APPLICANT: Kieke, James Alvin

; APPLICANT: Potter, David George

; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

; FILE REFERENCE: LEX-0144-USA

; CURRENT APPLICATION NUMBER: US/09/795,926

; CURRENT FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: US 60/185,920

; PRIOR FILING DATE: 2000-02-29

; PRIOR APPLICATION NUMBER: US 60/186,558

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 60/191,849

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 7

; LENGTH: 447

; TYPE: DNA

; ORGANISM: homo sapiens

US-09-795-926-7

Alignment Scores:

Pred. No.: 2,12e-09 Length: 447

Score: 155.50 Matches: 42

Percent Similarity: 45.74% Conservative: 17

Best Local Similarity: 32.56% Mismatches: 47

Query Match: 7.55% Indels: 23

DB: 4 Gaps: 5

US-09-854-122-16 (1-391) x US-09-795-926-7 (1-447)

Qy 82 AspAsnPhelysAlaArgGluThrAspIleLeuThrThrLeuProLysAlaGlyThr 101

Db 31 GACACCTTCGAAGCCAGACATGATGACATGCTGTAGCATCTTATCCAAAGTGGGTTC 90

Qy 102 ThrTrp-----ThrLysAlaLeuThrPheAlaLeuThrArgAspValAsn 117

Db 91 AACTGGATTCTCCACATGTCAGTGAATTAATATATGCTGTCTTAAAGAAAGTATAAA 150

Qy 118 HisProSerSerProThrHisProLeuLeuPheAsnProHisSerCysValGlnAsn 137

Db 151 TATCCAGAAATCCAGTT-----CTTGAATGTCGGGATTCA 186

Qy 138 LeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGlu 157

Db 187 GAAAAATAT-----CAGAGAAATGAAAGGCTTTCCA-----216

Qy 158 SerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLys 177

Db 217 TCACCAAGGATTTGGCACTCACTCCACTATGACAAATACCTGGGTCTATCTTCGAG 276

Qy 178 SerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheValSerPheTrp 197

Db 277 AATAAGCCAGATATTGGTGATATTTCCGAACCCATAAGATACAGCAGTATCTTTTGG 336

Qy 198 LysPheGlyAsnLeuLeuAsnProAsp 206

Db 337 CATTTCCACACGATGTC---CCCGAT 360

RESULT 14

US-09-795-926-11

; Sequence 11, Application US/09795926

; Patent No. 6555669

; GENERAL INFORMATION:

; APPLICANT: Donoho, Gregory

; APPLICANT: Hilbun, Erin

; APPLICANT: Turner, C. Alexander Jr.

; APPLICANT: Friedrich, Glenn

; APPLICANT: Abuin, Alejandro

; APPLICANT: Sambrowicz, Brian

; APPLICANT: Sands, Arthur T.

; APPLICANT: Walke, D. Wade

; APPLICANT: Wilganowski, Nathaniel L.

; APPLICANT: Hu, Yi

; APPLICANT: Kieke, James Alvin

; APPLICANT: Potter, David George

; TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE PROTEINS AND

; FILE REFERENCE: LEX-0144-USA

; CURRENT APPLICATION NUMBER: US/09/795,926

; CURRENT FILING DATE: 2001-02-28

; PRIOR APPLICATION NUMBER: US 60/185,920

; PRIOR FILING DATE: 2000-02-29

; PRIOR APPLICATION NUMBER: US 60/186,558

; PRIOR FILING DATE: 2000-03-02

; PRIOR APPLICATION NUMBER: US 60/191,849

; PRIOR FILING DATE: 2000-03-24

; NUMBER OF SEQ ID NOS: 47

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 561

; TYPE: DNA

; ORGANISM: homo sapiens

US-09-795-926-11

Alignment Scores:

Pred. No.: 3,05e-09 Length: 561

Score: 155.50 Matches: 42

Percent Similarity: 45.74% Conservative: 17

Best Local Similarity: 32.56% Mismatches: 47

Query Match: 7.55% Indels: 23

DB: 4 Gaps: 5

US-09-854-122-16 (1-391) x US-09-795-926-11 (1-561)

Qy 82 AspAsnPhelysAlaArgGluThrAspIleLeuThrThrLeuProLysAlaGlyThr 101

Db 145 GACACCTTCGAAGCCAGACATGATGACATGCTGTAGCATCTTATCCAAAGTGGGTTC 204

Qy 102 ThrTrp-----ThrLysAlaLeuThrPheAlaLeuThrArgAspValAsn 117

Db 205 AACTGGATTCTCCACATGTCAGTGAATTAATATATGCTGTCTTAAAGAAAGTATAAA 264

Qy 118 HisProSerSerProThrHisProLeuLeuPheAsnProHisSerCysValGlnAsn 137

Db 265 TATCCAGAAATCCAGTT-----CTTGAATGTCGGGATTCA 300

Qy 138 LeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGlu 157

Db 301 GAAAAATAT-----CAGAGAAATGAAAGGCTTTCCA-----330

Qy 158 SerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLys 177

Db 331 TCACCAAGGATTTGGCACTCACTCCACTATGACAAATACCTGGGTCTATCTTCGAG 390

Qy 178 SerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheValSerPheTrp 197

Db 391 AATAAGCCAGATATTGGTGATATTTCCGAACCCATAAGATACAGCAGTATCTTTTGG 450



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model  
Run on: March 20, 2004, 03:52:24 ; Search time 444 Seconds  
(without alignments)  
3257.588 Million cell updates/sec

Title: US-09-854-122-16  
Perfect score: 2059  
Sequence: 1 TRGITGIAVACLPLINAGIL.....CLLSYSFVNNFLLKKKK 391

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2438257 seqs, 1849576744 residues

Total number of hits satisfying chosen parameters: 4876514

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -p2n.model -DEV=xlh  
-Q=/cgn2\_1/USPTO.spool/US09854122/runat\_18032004/app.query.fasta\_1.583  
-DB=Published Applications NA -QFMT=fastap -SURFIX=rnpb -MINMATCH=0.1  
-LOEFLC=0 -LOEPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blomsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZ=500 -MINLEN=0  
-MAXLEN=200000000 -USER=US09854122 @CGN 1 1 221 @runat\_18032004 100001\_3284  
-NCPU=6 -ICPU=3 -NO MAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSPBLOCK=100  
-LONGLOG -DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5  
-FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DSLOP=6 -DELEX=7

Database : Published Applications NA:

1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq:\*  
2: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq:\*  
3: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq:\*  
4: /cgn2\_6/ptodata/1/pubpna/US06\_PUBCOMB.seq:\*  
5: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq:\*  
6: /cgn2\_6/ptodata/1/pubpna/PCTUS\_PUBCOMB.seq:\*  
7: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq:\*  
8: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq:\*  
9: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq:\*  
10: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq:\*  
11: /cgn2\_6/ptodata/1/pubpna/US09C\_PUBCOMB.seq:\*  
12: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq:\*  
13: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq:\*  
14: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq:\*  
15: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq:\*  
16: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq:\*  
17: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq:\*  
18: /cgn2\_6/ptodata/1/pubpna/US60\_PUBCOMB.seq:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
-----					

1	1996	96.9	1192	9	US-09-854-122-15	Sequence 15, Appl
2	589	28.6	1311	12	US-10-424-599-58971	Sequence 58971, A
3	545	26.5	996	9	US-09-938-842A-1027	Sequence 1027, Ap
4	545	26.5	996	11	US-09-938-842A-1027	Sequence 1027, Ap
5	537.5	26.1	981	9	US-09-938-842A-2049	Sequence 2049, Ap
6	537.5	26.1	981	11	US-09-938-842A-2049	Sequence 2049, Ap
7	527	25.6	271990	14	US-10-195-144-87	Sequence 87, Appl
8	527	25.6	271990	15	US-10-345-072-87	Sequence 87, Appl
9	525.5	25.5	1403	12	US-10-425-114-21478	Sequence 21478, A
10	525.5	25.5	1526	12	US-10-425-114-27795	Sequence 27795, A
11	521.5	25.3	1017	15	US-10-259-194A-3	Sequence 3, Appli
12	519.5	25.2	1047	15	US-10-259-194A-379	Sequence 379, App
13	501.5	24.4	984	14	US-10-259-165-443	Sequence 443, App
14	501.5	24.4	987	14	US-10-259-165-443	Sequence 443, App
15	483	23.5	2067	15	US-10-195-144-80	Sequence 80, Appl
16	483	23.5	2067	15	US-10-345-072-80	Sequence 80, Appl
17	479	23.3	1032	15	US-10-259-194A-329	Sequence 329, App
18	446.5	21.7	1041	14	US-10-259-165-331	Sequence 331, App
19	444.5	21.6	787	12	US-10-424-599-8242	Sequence 8242, A
20	443.5	21.5	1461	15	US-10-259-194A-131	Sequence 131, App
21	443.5	21.5	1461	15	US-10-260-238-1003	Sequence 1003, Ap
22	424	20.6	1050	15	US-10-259-194A-15	Sequence 15, Appl
23	321	15.6	515	10	US-09-770-961-120	Sequence 120, App
24	268	13.0	1004	13	US-10-199-330-3	Sequence 3, Appli
25	268	13.0	1004	14	US-10-199-334-3	Sequence 3, Appli
26	268	13.0	1004	14	US-10-199-329-3	Sequence 3, Appli
27	266.5	12.9	1167	12	US-10-072-012-205	Sequence 1, Appli
28	264	12.8	942	13	US-10-199-330-1	Sequence 1, Appli
29	264	12.8	942	14	US-10-199-334-1	Sequence 1, Appli
30	264	12.8	942	14	US-10-199-329-1	Sequence 1, Appli
31	260	12.6	922	13	US-10-199-330-2	Sequence 2, Appli
32	260	12.6	922	14	US-10-199-334-2	Sequence 2, Appli
33	260	12.6	922	14	US-10-199-329-2	Sequence 2, Appli
34	257	12.5	1363	9	US-09-917-800A-1391	Sequence 1391, Ap
35	257	12.5	1363	9	US-10-388-934-204	Sequence 204, App
36	255	12.3	956	15	US-10-116-275-232	Sequence 232, App
37	253.5	12.3	921	9	US-09-898-570-25	Sequence 25, Appl
38	253.5	12.3	921	10	US-09-839-446-25	Sequence 25, Appl
39	253	12.3	462	15	US-10-259-194A-643	Sequence 643, App
40	253	12.3	515	15	US-10-259-194A-492	Sequence 492, App
41	250	12.1	1000	15	US-10-388-934-493	Sequence 493, App
42	250	12.1	1000	15	US-10-388-934-789	Sequence 789, App
43	250	12.1	1000	15	US-10-388-934-825	Sequence 825, App
44	248.5	12.1	1229	9	US-09-816-790-19	Sequence 19, Appl
45	245.5	11.9	1138	9	US-09-981-353-100	Sequence 100, App

#### ALIGNMENTS

RESULT 1

US-09-854-122-15  
; Sequence 15, Application US/09854122  
; Patent No. US20020016980A1  
; GENERAL INFORMATION:  
; APPLICANT: SMITH, ROBERT  
; TITLE OF INVENTION: TRANSGENIC PLANTS INCORPORATING TRAIT OF ZOSTERA MARINA  
; FILE REFERENCE: PHA-007.01  
; CURRENT APPLICATION NUMBER: US/09/854,122  
; PRIOR FILING DATE: 2001-09-10  
; PRIOR FILING DATE: 2000-05-10  
; NUMBER OF SEQ IDS: 51  
; SOFTWARE: Patentin Ver. 2.1  
; SEQ ID NO 15

LENGTH: 1192  
TYPE: DNA  
ORGANISM: Zostera marina  
FEATURE:  
NAME/KEY: CDS  
LOCATION: (1..33, 37..1041, 1045..1059, 1063..1077,  
1081..1086, 1090..1119, 1123..1191)  
US-09-854-122-15









QY	57	Trp	---	Gly	Asp	Thr	Met	Arg	Leu	Tyr	Ile	ys	Gly	Phe	Trp	Gln	Met	Gly	Tyr	Leu	Val	Pro	75	
Db	97	TGG	T	TAG	T	AG	T	GA	A	A	T	A	T	A	T	A	T	A	T	A	T	A	156	
QY	76	Gly	I	le	Met	A	la	P	he	G	lu	Asp	Asn	P	he	L	ys	A	l	Arg	G	lu	Thr	95
Db	157	GGA	A	T	T	GA	T	CT	GC	A	A	A	A	A	A	A	A	A	A	A	A	A	216	
QY	96	Leu	P	ro	L	ys	A	l	a	g	l	y	Th	r	Th	r	P	ro	L	ys	A	l	a	115
Db	217	AAT	C	T	A	A	T	CAG	T	AC	C	A	C	T	T	G	T	T	A	A	A	A	273	
QY	116	Val	A	sn	H	i	s	-----	Pro	S	er	Pro	Th	r	H	i	s	P	ro	L	eu	Leu	P	131
Db	274	----	----	CAC	A	G	T	T	T	C	A	G	T	T	T	C	T	C	T	G	T	A	327	
QY	132	H	i	S	er	C	ys	V	al	G	lu	A	sn	L	eu	G	l	y	T	yr	T	yr	Met	151
Db	328	CAC	T	T	T	T	G	T	C	C	T	T	C	T	T	C	T	T	G	G	A	A	375	
QY	152	Leu	A	sp	Met	L	eu	A	n	-----	G	lu	S	er	Pro	Arg	L	eu	P	he	A	l	170	
Db	376	TTC	G	A	T	T	T	C	C	A	G	T	T	C	T	C	C	A	A	C	T	G	A	435
QY	171	Leu	P	ro	A	la	S	er	V	al	L	eu	L	ys	S	er	G	l	y	Th	r	L	ys	190
Db	436	CTC	C	C	C	C	G	A	G	T	C	T	T	G	A	A	G	T	C	G	T	T		

```

RESULT 6
US-09-938-842A-2049
; Sequence 2049, Application US/09938842A
; Publication No. US2004009476A9
; GENERAL INFORMATION:
; APPLICANT: Harper, Jeff
; APPLICANT: Kreps, Joel
; APPLICANT: Wang, Xun
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: STRESS-REGULATED GENES OF PLANTS, TRANSGENIC PLANTS CONTAINING
; TITLE OF INVENTION: SAME, AND METHODS OF USE
; FILE REFERENCE: SCRI1300-3

```

```

; CURRENT APPLICATION NUMBER: US/09/938,842A
; CURRENT FILING DATE: 2001-08-24
; PRIOR APPLICATION NUMBER: US 60/227,866
; PRIOR FILING DATE: 2000-08-24
; PRIOR APPLICATION NUMBER: US 60/264,647
; PRIOR FILING DATE: 2001-01-16
; PRIOR APPLICATION NUMBER: US 60/300,111
; PRIOR FILING DATE: 2001-06-22
; NUMBER OF SEQ ID NOS: 5379
; SEQ ID NO 2049
; LENGTH: 981
; TYPE: DNA
; ORGANISM: Arabidopsis thaliana
US-09-938-842A-2049

Alignment Scores:
Pred. No.: 1,97e-54 Length: 981
Score: 537.50 Matches: 123
Percent Similarity: 57.10% Conservative: 54
Best Local Similarity: 39.68% Mismatches: 112
Query Match: 26.10% Indels: 21
DB: 9 Gaps: 9

US-09-854-122-16 (1-381) x US-09-938-842A-2049 (1-981)

QY 37 AspSerLysMetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyr 56
DB 37 GATGAAGATCTGCACACAAAGAAACAAGAGCTCTGATCTCTCTCTCTTAAGAGAAAGGT 96
QY 57 TTP---GlyAspThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValPro 75
DB 97 TGGTTAGTGAAGTAAATATATGAATCCAAAGGACTTTGGCACACACAGACTATTTTACA 156
QY 76 GlyIleMetAlaPheGluAspAsnPheLysAlaArgGluThrAspIleIleLeuThrThr 95
DB 157 GGAATCTTGATCTGCCAAAAACGCTTTGAAGCTAAAGATTCGACATTATCTCGTCACT 216
QY 96 LeuProLysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAsp 115
DB 217 AATCTTAATCAGGTACCACCTTGGTTAAAGAGCTCTTGCTCTCTCTTAACCGA--- 273
QY 116 ValAsnHisL-----ProSerSerProThrHisProLeuLeuPhePheAsnPro 131
DB 274 -----CACAGTTTCAGATTTCTTCTCTGGTAACCATCTCTTCTGTGTACCAATCCA 327
QY 132 HisSerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAsp 151
DB 328 CACCTTCTGTGCCCTTCTTGAAGAGAGTTTACTAC-----GAGTCCCCCAGAT 375
QY 152 LeuAspMetLeuAsn---GluSerProArgLeuPheAlaGlyHisIleProTyrSerLeu 170
DB 376 TTCGATTTCTCCAGTTTGCCTTCTCCAAGACTGATGAACACGCACATATCGCATCTTTCG 435
QY 171 LeuProAlaSerValLeuLysSerGlyThrLysIleIleAsnIleSerArgAsnArgLys 190
DB 436 CTCGCCGAGTCTGTAAAGAGCTCGCTCTGTGAAGATTGTGTATTGTGTAGGAACCCATAAG 495
QY 191 SerThrPheValSerPheTrpLysPheGlyAsnLeuIleAsnProAspLysLeuLeuAsp 210
DB 496 GACATGTTTGTGCTTATGCAATTTTGGAAAAAGCTAGCTCTCTGAGGAACCCGCGAT 555
QY 211 -----LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGlu 228
DB 556 TATCTCATCGAAAAAGCGTTTGAAGCGCTTTTGTGAAGGGAAGTTTATAGTGTGACCCCTTT 615
QY 229 TTPAsnPheGlnAlaGluPheThrAsnAlaSerThrAsnSerAsn---LeuLeuLeu 247
DB 616 TGGATCATATATGGAGTACTGGTATGCAAGCGCGGAGNATCCGAACAGAGTCTTGTT 675
QY 248 LeuSerTyrGluMetLeuGluLysProValIgluAsnValLysLysLeuAlaGluPhe 267
DB 676 GTTACTTACGAGAGCTTAAGAACACACGCCAAGTTTGAGATGAAGCGGATCGCGAGTTTC 735

```

Qy 268 MetGlyCysGlyPheThrAspGluGluValGlnGlyLeuValAspGluLeuValys 287  
 Db 736 TTGAATGTCCTTATTGAGAGAGAA-----GTGAGAGATTGTGAAG 783  
 Qy 288 LeuCysSerPheAspLeuLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsn 307  
 Db 784 TTGTGTAGCTTTGAGAGTTTAAGTAATTTGGAAGTTTAAACAAGAGGAAATTTGCCAAAT 843  
 Qy 308 SerLysIleAspAsnLysHisPheArgGlyGluValArgAspTyrAlaAsnTyr 327  
 Db 844 GGA---ATAGAGACTAAACTTTCTTTAGAAAAGGAGAGATTGGAGGATGGAGATACT 900  
 Qy 328 LeuThrSerGluMetLysLysLysLeuGlu 337  
 Db 901 TTGAGTGAGTCATTGGCAGAGAAATTGAT 930

RESULT 7  
 US-10-195-144-87/c  
 ; Sequence 87, Application US/10195144  
 ; Publication No. US20030126646A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BROWN, GREGORY G.  
 ; APPLICANT: FORMANOVA, NATASA  
 ; APPLICANT: DENDY, CHARLES  
 ; APPLICANT: LANDREY, BENOIT S.  
 ; APPLICANT: CHEUNG, WING  
 ; APPLICANT: JIN, HUA  
 ; TITLE OF INVENTION: NUCLEAR FERTILITY RESTORER GENES AND METHODS OF USE IN  
 ; TITLE OF INVENTION: PLANTS  
 ; FILE REFERENCE: 16313-0136  
 ; CURRENT APPLICATION NUMBER: US/10195,144  
 ; PRIOR FILING DATE: 2002-10-01  
 ; PRIOR APPLICATION NUMBER: 60/305,026  
 ; PRIOR FILING DATE: 2001-07-12  
 ; PRIOR APPLICATION NUMBER: 60/305,363  
 ; PRIOR FILING DATE: 2001-07-13  
 ; PRIOR APPLICATION NUMBER: 60/308,736  
 ; PRIOR FILING DATE: 2001-07-30  
 ; NUMBER OF SEQ ID NOS: 128  
 ; SOFTWARE: Patent In Ver. 2.1  
 ; SEQ ID NO 87  
 ; LENGTH: 271990  
 ; TYPE: DNA  
 ; ORGANISM: Raphanus sativum  
 ; FEATURE:  
 ; NAME/KEY: modified base  
 ; LOCATION: (144241)..(144300)  
 ; OTHER INFORMATION: a, t, c, g, other or unknown  
 US-10-195-144-87

Alignment Scores:  
 Pred. No.: 5,34e-49 Length: 271990  
 Score: 527.00 Matches: 120  
 Percent Similarity: 57.81% Conservative: 65  
 Best Local Similarity: 37.50% Mismatches: 115  
 Query Match: 25.59% Indels: 21  
 DB: 14 Gaps: 8

US-09-854-122-16 (1-391) x US-10-195-144-87 (1-271990)

Qy 29 LysAsnGlnGluLysGluGlu-----AspSerLys 39  
 Db 244900 ARGAAACAAGAACAAACAAATGGATCCGAGGAGCTTCCTCGACACCTTAGACAGCAAG 244841

Qy 40 MetTyrLysArgTyrArgGluLeuValSerSerLeuProSerAsn---AspTyrTrpGly 58  
 Db 244840 ATAAGTGAAGAACCAAGAGGTGATCTCTTCCTTCCTTCACACACAGATTTACCAAGG 244781

Qy 59 AspThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyLeuMet 78  
 Db 244780 CATAAGATTTCGTAATATCAAGATGTTGATATATATACACACCTCCAGGTGTCTCA, 244721

Qy 79 AlaPheGluAspAsnPheLysAlaArgGluThrAspIleLeuThrThrLeuProLys 98

Db 244720 AT--TTCCAGAGAGGTTTTCACCGCAGACACTGATGTAATCATTTGCTTCGTACCCCAA 244662  
 Qy 99 AlaGlyThrThrTrpThrLysAlaLeuThrPheAlaLeuThrArgAspValAsnHis 118  
 Db 244661 TCAGGCACTACTTGGCTCAAGGCCCTCAGCTCGCTCTGTGAGAGATCAAGAACCAAC 244602  
 Qy 119 ProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeu 138  
 Db 244601 TCTTCTGAT-----CATCCTCTCTATATCATATCTCATGCGATATATACCATCTTG 244548  
 Qy 139 GluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSer 158  
 Db 244547 GAG---ATCGATGTGTATCCACGAAAGCTCAAGTCTCACTAGCAAGTTCTCAGCACCT 244491  
 Qy 159 ProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeuLysSer 178  
 Db 244490 CCGAGGCTGTTCTCGACTCATCATGCTGCACACGATCCACGAGCACTCAGCACTCT 244431  
 Qy 179 GlyThrLysIleAlaAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLys 198  
 Db 244430 CCTTGCAGATTTGTGTACGTGTGCAGAACCTGGAAGGACACGTTCATCTCTGTTGGTT 244371  
 Qy 199 PheGly-----AsnLeuIleAsnProAspLysLeuLeuAspLeuGluLysSerVal 215  
 Db 244370 TACAGCTGTGTATATATAAATCGAACCAACGAGAGTT---CTCGAGTCTATGTTT 244314  
 Qy 216 AspIlePheAlaSerGlyLysSerPheCysGlyProGluTrpAsnPheGluAlaGluPhe 235  
 Db 244313 AACGAGTCTCTCGATGGAAACCAACTATTTTGGACCTTTTGGGATCATCTCTTGAAGTTAC 244254  
 Qy 236 ThrAsnAlaIleSerThrAsn---SerAsnLeuLeuLeuSerTyrGluGluMetLeu 254  
 Db 244253 TCGAGAGAGAGCTTGGAGACCCCAAGCATGCTCTTTTCATGAGGTATGAGAGATGAAA 244194  
 Qy 255 GluLysProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAsp 274  
 Db 244193 GCGAGCCTCGTGATCAGATCAAGAGCTTCGGGACTTCTTGGGATGTCCTTTTACTAAG 244134  
 Qy 275 AspGluGluLysGlnGlyLeuValAspGluIleValLysLeuCysSerPheAspAsnLeu 294  
 Db 244133 CAAGAAGAAGATAGTGGATCTGTGCACGGATCTTGGACCTCTGCTCTCCGTAATCTG 244074  
 Qy 295 LysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHis 314  
 Db 244073 AGCAGTTTGGAGCTAAACAAACAGGACAAATAACAAT-----GTGAGCACAGTTT 244020  
 Qy 315 PhePheArgLysGlyGluValArgAspTyrAlaAsnTyrLeuThrSerGluMetIleLys 334  
 Db 244019 TTTTTCGTAAGGAGAGAGTGGTCACTCGAAAAATTTATCTTACGCTCTGAATGGAGAA 243960

RESULT 8  
 US-10-345-072-87/c  
 ; Sequence 87, Application US/10345072  
 ; Publication No. US20030237112A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: BROWN, GREGORY G.  
 ; APPLICANT: FORMANOVA, NATASA  
 ; APPLICANT: DENDY, CHARLES  
 ; APPLICANT: LANDREY, BENOIT S.  
 ; APPLICANT: CHEUNG, WING  
 ; APPLICANT: JIN, HUA  
 ; APPLICANT: LAI, FANG MING  
 ; APPLICANT: LEFOREST, MARTIN  
 ; TITLE OF INVENTION: NUCLEAR FERTILITY RESTORER GENES AND METHODS OF USE IN  
 ; TITLE OF INVENTION: PLANTS  
 ; FILE REFERENCE: 16313-0210  
 ; CURRENT APPLICATION NUMBER: US/10/345,072  
 ; CURRENT FILING DATE: 2003-01-16  
 ; PRIOR APPLICATION NUMBER: PCT/US02/22217  
 ; PRIOR FILING DATE: 2002-07-12  
 ; PRIOR APPLICATION NUMBER: 60/305,026  
 ; PRIOR FILING DATE: 2001-07-12



```
Db 424 CCGCCATTGGATGGCGACCATCCGCTCGTGTGCAACCCCTCATGACTGTGTGAGGTTTC 483
Qy 138 LeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGlu 157
Db 484 CTCGATGCTAATTTCAACCCAGCAGAGATGAGTGGAGGCTCTC-----CCG 531
Qy 158 SerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeu--- 176
Db 532 TCCCGCGCGGTGTAGCCACGACGCTCCCTCTACTCCCTGTGCGCGCAGCATCACCGGG 591
Qy 177 -----LysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPhe 193
Db 592 GATAGGAGGCCCTCGGGTGGCGATGCTGTATGCTGCGCGGNAACCCAGAGCGCGCTG 651
Qy 194 ValSerPheThrLysPhe-----GlyAsnLeuLeuAsnProAsp---LysLeu 208
Db 652 GTCTCTGCTGCTGCTTCCAGGAGGAGCGCGCTCGCGCTGGGGGTGACGCTCGATCG 711
Qy 209 LeuAspLeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGlu 228
Db 712 TTCAGATCCAGGAGGCTTGGAGCTTTCTGATGCTGCTGTATGTCGGCGCCGCGAG 771
Qy 229 TrpAsnPheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeu 247
Db 772 TGGGAGCAGCTCCTCCAGTACTGGGAGGAGGAGCGGTGAGGAGCGCTGACAGGGTGTTC 831
Qy 248 LeuSerTyrGluGluMetLeuLysProValGluAsnValLysLysLeuAlaGluPhe 267
Db 832 CTTCGGTACGAGGAGATGCTGATTGATCCCGAGGACACGTCAGAGAGCTCGCCAAAGTTC 891
Qy 268 MetGlyCysGlyPheThrAspAspGluLysGlnGlyIleValAspGluIleValLys 287
Db 892 ATGGGGTGTGATTTTCTGAGGAGGAGGAGGAGCAGCGGGTGTGTGAGCGGCATCGTGGAG 951
Qy 288 LeuCysSerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSerTyrAsn 307
Db 952 CTGTGTACCTCGGCAAGATGAGGACATGAGGAGTGAACAGAAATGGAGCAATATGTTG 1011
Qy 308 SerLysIleAspAsnLysHisPhePheArgLysGlyGluValArgAspTrpAlaAsnTyr 327
Db 1012 GGG---GTCAGAACAGAAAGCTACTTTCAGGAGGAGGAGTGTGCGGACTGAGCAACCAT 1068
Qy 328 LeuThrSerGluMetIleLysLysLeuGlu-----Thr 338
Db 1069 ATGACGCGGACATGCGCAGAGGCTAGACAAAGTCTGATAGAGATGCAATGCAAGGACT 1128
Qy 339 AlaGlyLysIleAsnGluSer-----GluLys 347
Db 1129 GGATT-CAGCTTACAGCAAGCATGATGACCACTCGGCTCATGTTAGTGGAACTC 1187
Qy 348 HisLeuLeuSerAsnLysAsnLeuThrAsnPheAsnLeu----- 360
Db 1188 CATAATGCCAAGAACAACTCGTTTGTGCGTATTCGTTGATGGATGGAGATCATACCAGC 1247
Qy 361 -----AsnAsnTyrCysGluLysSerAsnIleSerLeuLysLeuSer 375
Db 1248 CAGCCATGTGAGCATCTCTTTTGTGATGAGGAGGAAATGATTTCGCTATATTCAT 1307
Qy 376 ---TyrSerPheValIleAsnAsn 382
Db 1308 CTGTATAACTTTGTTACAAACCAT 1331

RESULT 10
US-10-425-114-27795
; Sequence 27795, Application US/10425114
; Publication No. US20040034868A1
; GENERAL INFORMATION:
; APPLICANT: Liu, Jingdong
; APPLICANT: Zhou, Yihua
; APPLICANT: Kovalic, David K.
; APPLICANT: Screen, Steven E
; APPLICANT: Tabaska, Jack E
```

```
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated With
; FILE OF INVENTION: Plants and Uses Thereof for Plant Improvement
; FILE REFERENCE: 38-21(53313)B
; CURRENT APPLICATION NUMBER: US/10/425,114
; NUMBER OF SEQ ID NOS: 2003-04-28
; NUMBER OF SEQ ID NOS: 73128
; SEQ ID NO 27795
; LENGTH: 1526
; TYPE: DNA
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: LIB4740-009-D3_FLI
US-10-425-114-27795
```

```
Alignment Scores:
Pred. No.: 1,178-52 Length: 1526
Score: 525.50 Matches: 129
Percent Similarity: 50.49% Conservative: 77
Best Local Similarity: 31.62% Mismatches: 134
Query Match: 25.52% Indels: 69
DB: 12 Gaps: 15
```

US-09-854-122-16 (1-391) x US-10-425-114-27795 (1-1526)

```
Qy 30 AsnGluGlnLysGluGluAsp-----SerLysMetTyrLysArgTyrArgGluIle 47
Db 231 AATGCACAAAGACGACGATCACATTGGCGAGTCTCTCTCACTCCAAACATAGCCAGCATC 290
Qy 48 ValSerSerLeuProSerAsnAspTyrTrpGlyAspThr-----Met 61
Db 291 ATCCCTTCCTCCCTCCTC-----GAGACCAGGTGCCACCGTTCCTCTCTC 335
Qy 62 ArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeu-----ValProGlyIle 77
Db 336 CGCGCGTACGCAACTTTTGGGTGCTGAGGTATATTGAAGGCGAGACCTCCCGCGCAT 395
Qy 78 MetAlaPheGluAspAsnPheValaArgGluThrAspIleIleLeuThrThrLeuPro 97
Db 396 -----CATTCCTGCTTCAAGCGGAGACCCACCGAGCTTGTGTCGCAAGCTTCCCG 446
Qy 98 LysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsn 117
Db 447 AAGTCTGGCACCATTTGGCTCAAGGCCCTTGCTTCGCGACGCTGAAGCGTTCACGAT 506
Qy 118 HisProSerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsn 137
Db 507 CCGCCATTGGATGGCGACCATCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 566
Qy 138 LeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGlu 157
Db 567 CTCGATGCTAATTTCAACCCAGCAGAGGATGAGTTGGAGGCTCTC-----CCG 614
Qy 158 SerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeu--- 176
Db 615 TCCCCGCGGTGTAGCCACGACCTCCCTCTACTCCCTGCTGCTGCTGCTGCTGCTGCTGCTG 674
Qy 177 -----LysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPhe 193
Db 675 GATAGGAGCGCTCGGGTGGCGATGCTGTGCTGCGCGGAAACCCAGAGCGCGCTG 734
Qy 194 ValSerPheThrLysPhe-----GlyAsnLeuLeuAsnProAsp---LysLeu 208
Db 735 GTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 794
Qy 209 LeuAspLeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGlu 228
Db 795 TTCAGATCCAGGAGGCTTGGAGCTTTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 854
Qy 229 TrpAsnPheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeu 247
Db 855 TGGGAGCAGCTCTCCAGTACTGCGGAGGAGGAGCGGTGAGGAGCGCTGAGGGGTGTTGTTTC 914
```

```
QY 248 LeuSerTyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGluPhe 267
Db 915 CTTGGTACGAGAGATGCTGATTGATCCCGAGGACACAGCTCAGGAAGTCCCAAGTTC 974
QY 268 MetGlyCysGlyPheThrAspGluGluLysGlnGlyValAspGluLysValLys 287
Db 975 ATGGGTGTGGATTCTCAGGAGGAGGAGGACGCGGGTGTGAGCGCCATCGTGGAG 1034
QY 288 LeuCysSerPheAspAsnLysLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsn 307
Db 1035 CTGTGTAGCTGGGCAAGATGAGGACATCGAGGTGAACAGAAATCGAAGCAATATGTTG 1094
QY 308 SerLysIleAspAsnLysHisPhePheArgLysGlyGluValArgAspTrpAlaAsnTyr 327
Db 1095 GGG--GTCAGAACAGAACTACTTCAGAGAGGAGTTCTGGGACTGGAGCAACCAT 1151
QY 328 LeuThrSerGluMetIleLysLysLeuGlu-----Thr 338
Db 1152 ATGACCGCGACATGCGCAGAGGCTAGACAAGGTCTAGAGGATGCAATTGCAAGGGACT 1211
QY 339 AlaGlyLysIleAsnGluSer-----GluLys 347
Db 1212 GATT--CAGCTTTACGACAAACATGATGATGACCTCGCTCATGTTGCGAATC 1270
QY 348 HisLeuLeuSerAsnLysAsnLeuThrAsnPheAsnLeu----- 360
Db 1271 CATAATGCCAAGAACAACTCGTTTCTGCGTATTGCTTGTGATGAGATCATACCAGC 1330
QY 361 -----AsnAsnTyrCysGluLysSerAsnIleSerLeuLysLeuSer 375
Db 1331 CAGCCATGGTGAGCATCTCTCTTTGTGATGAAGGGGAAATGATTGCGTAAATTCAT 1390
QY 376 ---TyrSerPheValIleAsnAsn 382
Db 1391 CTGTATAACTTTGTTACAAACCAT 1414
```

## RESULT 11

```
US-10-259-194A-3
; Sequence 3, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghassemanian, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Krepis, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Ricke, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 3
; LENGTH: 1017
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-194A-3
```

Alignment Scores:

Pred. No.: 1.79e-52

Length: 1017

1017

RESULT 12

```
Score: 521.50 Matches: 118
Percent Similarity: 56.03% Conservative: 54
Best Local Similarity: 38.44% Mismatches: 116
Query Match: 25.33% Indels: 19
DB: 15 Gaps: 8
US-09-854-122-16 (1-391) x US-10-259-194A-3 (1-1017)
QY 42 LysArgTyrArgGluLeuValSerSerLeuProSerAsnAspTyrTrpGlyAspThrMet 61
Db 97 GAGGAGTATCCCGCGCGTGTCCAGCCTCCCGAGC-----TATCCCAAGCTCGCGCTG 150
QY 62 ArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAlaPheGlu 81
Db 151 CGGCACCTACAGGCGCATGTGGCTGATGAGTACACCCCTCCAGGATCATGGCCATCCAG 210
QY 82 AspAsnPhe---LysAlaArgGluThrAspIleIleLeuThrThrLeuProLysAlaGly 100
Db 211 CGGAGCTTCGTGCGCGCCCGCCATGGCGAGCTCGTCTCGCGAGCCCGGTAAGTGGCGC 270
QY 101 ThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspValAsnHisProSer 120
Db 271 ACCACTTGGCTCAAGGCGCTCGCTTCGCGCTTGGCGCGCGCGGCTACTCCCTGCC 330
QY 121 SerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyr 140
Db 331 AGTGACCGGATCCGCTCCTCCGCTTTAACCCGACGACTGCGTCCCGTTCATGGAGGGG 390
QY 141 Leu-----TyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsn 155
Db 391 CGCATCTCTGAGGATGGGTGTAGATTGATGAGCTGCGG----- 432
QY 157 GluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerValLeu 176
Db 433 ---TCGCGGAGGCTCATGTCCACCCATGCAGCAGCTGCTTACCCAAAGTCCATGCA 489
QY 177 ---LysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThrPheValSer 195
Db 490 GATGAACCTGGCTCAAGGTGTATTATTTCAGGAGGCGGAGGATATCTCTGTTCA 549
QY 196 PheTrpLysPheGlyAsnLeuLeuAsnProAspLysLeuLeuAspLeuLysSerVal 215
Db 550 GCCTGGCACTTCTTCAGAAATTATAGAGCTGAT-----CTTTCATTCAAGAAGTGTTC 603
QY 216 AspIlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPhe 235
Db 604 GAGGCTGCTCGCATGGCAAGTCTTTCACCGCGCTATCTGGGATCACAATTATGGCTAC 663
QY 236 ThrAsnAlaAlaSerThrAsn---SerAsnLeuLeuLeuSerTyrGluGluMetLeu 254
Db 664 TGGATGCTTGCAGGCGGCAACCCAGAGAGGTCTCTGTTCTGTCTATGAGACCTTCTG 723
QY 255 GluLysProValGluAsnValLysIleValAlaGluPheMetGlyCysGlyPheThrAsp 274
Db 724 CGAGATCCGCGCAACATCGTCAGGAAGCTCCCGACTTCTTGGCCAGCCGCTTCTCAFCG 783
QY 275 AspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPheAspAsnLeu 294
Db 784 ACCGAGGAAGAGCTGGACTGTCAGATATTGAGGCTGTGCAGCTTCGAGAACCTG 843
QY 295 LysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLysIleAspAsnLysHis 314
Db 844 AAGAGCTTGGAGGTGAATAAGATGGGAGAGCA---TCGTTTCGCTTCCCGAAATCGGTCA 900
QY 315 PhePheArgLysGlyGluValArgAspTrpAlaAsnTyrLeuThrSerGluMetIleLys 334
Db 901 TACTTCCGGAAGGAGGCGGAGACTGGAAATCCACATGACCGCGGAAATGTTGGAG 960
QY 335 LysLeuGluThrAlaGlyLys 341
Db 961 TGTGTTGACCATCGTCAAG 981
```

```
US-10-259-194A-379
; Sequence 379, Application US/10259194A
; Publication No. US20040010815A1
; GENERAL INFORMATION:
; APPLICANT: Lange, Markus B.
; APPLICANT: Ghasseman, Majid
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Rickes, Darrell
; APPLICANT: Zhu, Tong
; TITLE OF INVENTION: IDENTIFICATION AND CHARACTERIZATION OF PLANT GENES
; FILE REFERENCE: 70029-NP
; CURRENT APPLICATION NUMBER: US/10/259,194A
; PRIOR FILING DATE: 2003-01-07
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 60/370,743
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; NUMBER OF SEQ ID NOS: 662
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 379
; LENGTH: 1047
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-194A-379

Alignment Scores:
Pred. No.: 3,27e-52 Length: 1047
Score: 519.50 Matches: 120
Percent Similarity: 56.38% Conservative: 48
Best Local Similarity: 40.27% Mismatches: 117
Query Match: 25.23% Indels: 13
DB: 8 Gaps: 5

US-09-854-122-16 (1-391) x US-10-259-194A-379 (1-1047)
QY 48 ValSerLeuProSerAsnAspTrpTrpGlyAspThrMetArgLeuTyLysGlyPhe 67
Db 130 GTCCTCGCCCTCGCGGGCGGTCTCTACGCGCCAGCGCATGGCTGCTACGGGGGACC 189
QY 65 TrpGlnMetGlyTyLeuValProGlyLeuMetAlaPhe---GluAspAsnPhelysAla 86
Db 190 TGGGTGTTCAGAGCTGGCGCGGCGGTGGTGGCCATGACCGCGGGCGGCTGTGGCC 249
QY 87 ArgGluThrAspIleLeuThrThrLeuProLysAlaGlyThrThrThrLysAla 106
Db 250 CGCGCGCGGCGAGTCTCTCTCCAGCTCCCAAGTCGGGACGACGTGGCTCAAGGCA 309
QY 107 LeuThrPheAlaLeuThrArgAspVal---AsnHisProSerSerProThrHisPro 125
Db 310 CTCGGTTCGCCACCATCGCGCGCGGTGGTGGCCCGCGCGCGCGGACCGACCCG 369
QY 126 LeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGlyLeuTyLys---MetGly 144
Db 370 CTCGCGCGCTAAACCGGACGACGTGGTGGCCCTCTCGACAGGCTCTTCGCGCGCGC 429
QY 145 ArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPheAlaGly 164
Db 430 CGCGACGCGCTTCTCGACGAGTG-----CCGTGCGCGGAGGCTCATGTGCACG 477
QY 165 HisLeuProTySerLeuLeuProAlaSerValLeu-----LysSerGlyThrLysIle 182
Db 478 CACATCGCGCTCTCGCTGTGGCGCGGGGTGGTGGCGGCACTCCACACCAAGATC 537
QY 183 IleAsnIleSerArgAsnArgLysSerThrPheValSerPheThrLysPheGlyAsnLeu 202
```

```
538 ATCTACATTTTCGCGGACCAAGAAAGATAGATTGGTCTCCATCTCGCACTTTCGGAACGC 597
203 ILeAsnProAspLysLeuLeuAspLeuLysSerValAspIlePheAlaSerGlyIle 222
598 AATGTGCCAGATCTGTTG-----CTCCAGAGTGTATGATCTATATGTGTGTCACCT 651
223 SerPheCysGlyProGluTTPAsnPheGlnAlaGluPheThrAsnAlaAlaSerThrAsn 242
652 GGGTTTGTGTCCTCTCTGGGATCACTGGTGGTGTACTGGAGGGGCAAGCAAAATAGAT 711
243 ---SerAsnLeuLeuLeuSerTyrgluGluMetLeuGluLysProValGluAsnVal 261
712 CTAGCCCGAGTCTCTTCTTGAATATGAGGAGGTCTTCGAGACCCAGTGAACACCGTT 771
262 LysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGluLysGlyIle 281
772 AGAGAGCTCGCACAAATTTGTCGGCGCCCATCTCTGACACAGAGGAGAGAGCTGGCATT 831
282 ValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAsnGlnValAsnLys 301
832 GTTGGGAGATCGTCAAGCTATGATAGAGAGTCTAAGAGCCAGCAAAAGCCACAAA 891
302 AsnGly---SerSerTyAsnSerLysIleAspAsnLysHisPheArgLysGlyGlu 320
892 GAAGGCATACAGAGTGTGTACATCAAGTCTCGCAGCATCTGTATTTAGGAAGGGGTG 951
321 ValArgAspTrpAlaAsnTyLeuThrSerGluMetIleLysLysLeuGluThr 338
952 GAAGGGGACTGAGGAAACCATGACTCTTAAGATGGGCGGACATCTGGATTGCG 1005

RESULT 13
US-10-259-165-443
; Sequence 443, Application US/10259165
; Publication No. US20030135888A1
; GENERAL INFORMATION:
; APPLICANT: Zhu, Tong
; APPLICANT: Wang, Xun
; APPLICANT: Chang, Hur-song
; APPLICANT: Briggs, Steven P.
; APPLICANT: Cooper, Bret
; APPLICANT: Glazebrook, Jane
; APPLICANT: Goff, Stephen A.
; APPLICANT: Katagiri, Fumiyaki
; APPLICANT: Kreps, Joel
; APPLICANT: Moughamer, Todd
; APPLICANT: Provart, Nicholas
; APPLICANT: Rickes, Darrell
; TITLE OF INVENTION: GENES THAT ARE MODULATED BY POSTTRANSCRIPTIONAL GENE SILENCING
; FILE REFERENCE: 70030-NP
; CURRENT APPLICATION NUMBER: US/10/259,165
; CURRENT FILING DATE: 2002-09-26
; PRIOR APPLICATION NUMBER: US 60/370,620
; PRIOR FILING DATE: 2002-04-04
; PRIOR APPLICATION NUMBER: US 60/368,327
; PRIOR FILING DATE: 2002-03-27
; PRIOR APPLICATION NUMBER: US 60/325,277
; PRIOR FILING DATE: 2001-09-26
; NUMBER OF SEQ ID NOS: 782
; SOFTWARE: PatentList.pl version 3.0.4 (C) 2001 Syngenta
; SEQ ID NO 443
; LENGTH: 984
; TYPE: DNA
; ORGANISM: Oryza sativa
US-10-259-165-443

Alignment Scores:
Pred. No.: 4.39e-50 Length: 984
Score: 501.50 Matches: 119
Percent Similarity: 50.80% Conservative: 39
Best Local Similarity: 38.26% Mismatches: 106
Query Match: 24.36% Indels: 47
DB: 14 Gaps: 8
```









GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:36:53 ; Search time 17 Seconds  
(without alignments)  
2212.405 Million cell updates/sec

Title: US-09-854-122-16  
Perfect score: 2059  
Sequence: 1 FRTGIGIAVACLPLIAGL.....CLLSYSFVNNFILKKKK 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_78.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	616	29.9	329	2 T47448	sulfotransferase-1
2	602	29.2	323	2 T47447	sulfotransferase-1
3	567	27.5	346	2 E86319	probable flavonol
4	562.5	27.3	338	2 A96769	protein flavonol s
5	558.5	27.1	350	2 H96768	protein flavonol s
6	546	26.5	325	2 T07833	probable steroid s
7	545	26.5	331	2 E84451	probable steroid s
8	544.5	26.4	326	2 F86407	probable sulfotran
9	537.5	26.1	326	2 A84452	probable steroid s
10	529.5	25.7	323	2 T07831	probable steroid s
11	527.5	25.6	302	2 S69188	probable flavonol
12	527	25.6	324	2 B84452	probable steroid s
13	525.5	25.5	333	2 A84523	probable steroid s
14	520.5	24.8	320	2 A40216	flavonol 4'-sulfot
15	500.5	24.3	324	2 T07832	probable steroid s
16	486.5	23.6	314	2 T06012	hypothetical prote
17	455	22.1	311	2 B40216	flavonol 3'-sulfot
18	309	15.0	295	2 S29045	amine sulfotransf
19	278.5	13.5	301	2 J00078	estrogen sulfotran
20	258	12.5	299	2 J01966	N-hydroxyvarylamine
21	255.5	12.4	304	2 A49098	estrogen sulfotran
22	255	12.4	294	2 J02229	aryl sulfotransfer
23	253.5	12.3	298	2 S28183	estrogen sulfotran
24	253	12.3	295	2 A41930	cytosolic sulfotra
25	252	12.2	296	2 J07921	estrogen sulfotran
26	250	12.1	295	2 I56606	estrogen sulfotran
27	250	12.1	295	2 I73679	estrogen sulfotran
28	245.5	11.9	295	2 I53266	testis-specific es
29	245	11.9	291	2 S10329	aryl sulfotransfer

30	243.5	11.8	285	1 I38548	alcohol sulfotrans
31	243.5	11.8	296	2 J05885	thyroid hormone su
32	239.5	11.6	286	2 A44011	adrenocortical est
33	234.5	11.4	286	2 J00152	alcohol sulfotrans
34	233	11.3	296	2 J01282	hydroxyvarylamine s
35	232	11.3	295	2 J00197	phenol sulfotransf
36	232	11.3	299	2 J05884	thyroid hormone su
37	229.5	11.1	295	2 J02523	aryl sulfotransfer
38	229	11.1	296	2 J07283	hydroxyvarylamine s
39	228.5	11.1	294	2 J05000	aryl sulfotransfer
40	226.5	11.0	295	2 G01843	aryl sulfotransfer
41	225.5	11.0	295	2 G01843	aryl sulfotransfer
42	225	10.9	295	2 S52791	aryl sulfotransfer
43	223.5	10.9	295	2 S52399	aryl sulfotransfer
44	223.5	10.9	295	2 J05248	aryl sulfotransfer
45	223.5	10.9	295	2 S52794	aryl sulfotransfer

ALIGNMENTS

RESULT 1

T47448  
sulfotransferase-like protein - Arabidopsis thaliana  
N:Alternate names: Protein T14D3.20  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 02-Sep-2000  
C:Accession: T47448  
R:Jordan, N.; Bangert, S.; Wiedelmann, R.; Voss, H.; Unseld, M.; Mewes, H.W.; Lemcke, K.  
Submitted to the Protein Sequence Database, February 2000  
A:Reference number: 224467  
A:Accession: T47448  
A:Status: Preliminary  
A:Molecule type: DNA  
A:Residues: 1-329 <J0R>  
A:Cross-references: EMBL:AL138649  
A:Experimental source: cultivar Columbia; BAC clone T14D3  
C:Genetics:  
A:Map position: 3  
A:Note: T14D3.20  
C:Superfamily: alcohol sulfotransferase

Query Match	29.9%	Score 616;	DB 2;	Length 329;
Best Local Similarity	41.2%	Pred. No. 1.3e-36;		
Matches 132;	Conservative 62;	Mismatches 100;	Indels 26;	Gaps 9;
QY	29	KNEQEKEDSNKYRYREIVSSLPN-DYWGDTWRLYKGFQWQGYLYPGIMAFEDNFKAR	87	
DB	11	RNDELSEES-----KTLISLSPSDKNSTGVNVCKYQGCWYTPPIQGVLFNFQNFKPQ	63	
QY	88	ETDIITLTPKAGTTWKALTFAILTRDVNHP-SPTHPLFFNPHSCVQNL-ELYMGR	145	
DB	64	DTDIIVASFFKCGITWLKALTFAIVRRS-KHPSHDDHPLSDNPHVLSPLSLEYLYLCS	122	
QY	146	ENTMPDLMLESPLPAGHTPYSLLPASVLKSGTKIINISRNKSTFVSFKF-----	199	
DB	123	EN--PDLTKFSSSLRFLSTHPSHTLOEGLKGSTCKIVMSRNVDLTLSVWHFCKQT	180	
QY	200	-GNLNPDKLIDLEKSDVIFASGISFCGPEWFOAEFTNAASTNSN-ILLLSYSEMLEKP	257	
DB	181	DNIIIS-----SVETTFEMFCRGVNFQFFWDHVLVSWRGSLSDPNHVLPMFKFEEMKEP	235	
QY	258	VENVKLAEPFGCGTDDERKQGIIVDEIKVLCSPDNLNKQOVKNKGSSYNKIDNKHFR	317	
DB	236	REQIKRLAEFLGCLFTKEEESGLVDEIIDLCRLNLSLEINKTKLHSTGRENKTFPR	295	
QY	318	KGEVDMANLYLTSEMIKKLE	337	
DB	296	KGEVGDWKNLYLTPEMENKID	315	

RESULT 2  
T47447

## RESULT 4

A96769 protein flavonol sulfoxtransferase F2P9\_3 [imported] - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C/Accession: A96769  
R/Theologians, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Cresay, T.H.;  
ansen, N.F.; Hughes, B.; Huizar, L.  
Nature 408, 816-820, 2000  
A/Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.;  
Rizzo, M.H.; Rooney, T.; Rowley, D.; Sakano, H.  
A/Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A/Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A/Reference number: A86141; MUID:21016719; PMID:111430712

A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-338 <STO>  
A:Cross-references: GB:AE005173; NID:G7109462; PIDX:AAF36726.1; GSPDB:GN00141  
C:Genetics:  
A:Gene: F3P9.3  
A:Map position: 1  
C:Superfamily: alcohol sulfotransferase

Query Match	27.3%;	Score	562.5;	DB	2;	Length	338;
Best Local Similarity	38.4%;	Pred. No.	8.9e-33;				
Matches	129;	Conservative	59;	Mismatches	123;	Indels	25;
Gaps	10;						
QY	33	EKEDSGMKYKRYREIVSSIPSDNYH--	GDWRLYKGFQMGLYVPGIMAFEDNFKARETD	90			
DB	14	ELTEFEKTOKYQDFIATUPLKSGWRPDEIIT	QYCGHHWQSCLEGLFHKDHFPEARPTD	73			
QY	91	IIILTLPKAGTWTTKALTAIILTRDWNHPSS	THPLLEFNPHSCVQNLEYLYMGRENTMP	150			
DB	74	FLVCSVPKGTGTTWLKALTAIVNR--	SRYDDAANELLKRNPHFPYVVEIDFA---FYP	122			
QY	151	DLIDMLNESPR-IPAGHIPPYSLLPASVLKSG	TGKTKIINISNRKSTFVFSFWKFGNLINPDK--	201			
DB	128	TVDVLQDRKNPLFSHTHINGLLPDSIVNSG	CKWVYIWDKPDKDTETISMWTF---LHKESQ	187			

Db 185 EGQLASLDSDFMFCGLSVYGYLHDVLGYWKAYQENPDRIFLFLRYETMRANPLPFVKR 24

QY 264 LAEPNGCGTDDDEKQGIQVDEIVKLCSDNFKQVNVKNGSSYNSK---IDNKHFRKGE 320  
DB 245 LAEPNGGCGTDEEENGVAEKVKVLCSDNFKQVNVKNGSSYNSK---IDNKHFRKGE 304  
QY 321 VRDWAYLTSEMIMKLETAGKINESEKH--LLSNKN 354  
DB 305 VGDWAYLTPEMAARID--GLVEKFKDTGLLQHDN 338

RESULT 5  
H96768  
protein flavonol sulfotransferase F2P9.4 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C:Accession: H96768  
R:Theologis, A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
ansen, N.F.; Hughes, B.; Hulzar, L.  
Nature 408, 815-820, 2000  
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziani,  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,  
ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
A:Reference number: A86141; MUID:21016719; PMID:11130712  
A:Accession: H96768  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-350 <STO>  
A:Cross-references: GB:AB005173; NID:G7109463; PIDN:AAF36727.1; GSPDB:GN00141  
C:Genetics:  
A:Map position: 1  
C:Superfamily: alcohol sulfotransferase

Query Match 27.1%; Score 558.5; DB 2; Length 350;  
Best Local Similarity 39.0%; Pred. No. 1.8e-32;  
Matches 129; Conservative 59; Mismatches 120; Indels 23; Gaps 9;

QY 29 KNEQKEEDSKWYKRYREIVSSLPNDYWGDTMEL--YKGFQMGYLVPGIMAFENFKA 86  
DB 22 ETKTSTEFKQKRYQDLSTFPHEKGWRPELEIYGGYWLPLSLLEGCIHAQEFFQA 81  
QY 87 RETDILTLPKAGTTWKALTFAILTRDVNHPSSPHLLFPNPHSCVNLEYLYNGRE 146  
DB 82 RPSDFLVCSYKPTGTTWLKALTFANRSRFDSS--NPLLRNPFVFPYIIDEF--- 135  
QY 147 NTMPDMLNESPRLPAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLINP 205  
DB 136 PFFPEVDVLKGNLTFSTHPIYELLSDSVKSKRVYIWRPKDTFISWTF---LHK 192  
QY 206 DK-----LLDLEKSVDFIPASGISFCGPEWNPQAEFTNAASTN--SNLLLSYEEMLEKPE 259  
DB 193 ERTLEGLPVSNLESFDMFCRLSGYGPYLNHILAYKAYQENPDRIILFLKYETWRADPLP 252  
QY 260 NVKLAEPWCGFTDDEKQGIQVDEIVKLCSDNFKQVNVKNGSSYNSK---IDNKHFF 316  
DB 253 YVKSIAEPWCGFTDDEKQGIQVDEIVKLCSDNFKQVNVKNGSSYNSK---IDNKHFF 312  
QY 317 RKGEVRDWAYLTSEMIMKLETAGKINESEK 347  
DB 313 RKGVGDWSNLYLTPEMAARID--GLMEKFK 341

RESULT 6  
T07833  
probable steroid sulfotransferase (EC 2.8.2.15) 3 - rape  
C:Species: Brassica napus (rape)  
C:Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 08-Oct-1999  
C:Accession: T07833  
R:Richard, M.; Nicolle, L.; Varin, L.  
submitted to the EMBL Data Library, April 1997

A:Reference number: Z16161  
A:Accession: T07833  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-325 <RIC>  
A:Cross-references: EMBL:AF000307; NID:G3420007; PIDN:AAC63113.1; PID:G3420008  
C:Genetics:  
A:Gene: ST3  
C:Superfamily: alcohol sulfotransferase  
C:Keywords: sulfotransferase

Query Match 26.5%; Score 546; DB 2; Length 325;  
Best Local Similarity 40.2%; Pred. No. 1.3e-31;  
Matches 123; Conservative 52; Mismatches 117; Indels 14; Gaps 7;

QY 37 DSKMYKRYREIVSSLPNDYWGDTMRLYKGFQMGYLVPGIMAFENFKAETDILIT 95  
DB 13 DEKLTQETRDILSSPSKGLVDSIOYQFQGRWHTALLOGLITTCQKHFKAOSDILIT 72  
QY 96 LPKAGTTWKALTFAILTRDVNHPSSPHLLFPNPHSCVNLEYLYMGRENTMPDLDL- 154  
DB 73 NPKSGTTLKLSLVFALINRHKFPVSSGDHPLLVNTNPHLLVPFMEGVY---ESPDFDS 128  
QY 155 LNESRRLPAGHIPYSLLPASVLKSGTKIINISRNKSTFVSFWKFGNLINPKLLD--LE 212  
DB 129 LLPFRLMNTHTSHLSLSPESVKSQCIIVCCRNPKDMFVSLMWHFGKKLAPQETADYPLE 188  
QY 213 KSVDFIFASGISFCGPEWNPQAEFTNAASTN--LILLSYEEMLEKPEVNVKLAEPWCG 271  
DB 189 KAVEAFQCGKTAGFPDHWLVYASLENPKNVLFVYBELKQTEVEVKRIAEFIGG 248  
QY 272 FTDDSEKQGIQVDEIVKLCSDNFKQVNVKNGSSYNSKIDNKHFFRKGEVDWNYLTSE 331  
DB 249 FTABEE---YSEIVKLCGSFESLSRLVNRQKLPNG-IETNAPFRKEIGGWRDTLS 303  
QY 332 MIKLE 337  
DB 304 LADAID 309

RESULT 7  
H84451  
probable steroid sulfotransferase [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C:Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 02-Mar-2001  
C:Accession: H84451  
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;  
M.; Koo, H.; Mofatt, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.  
eues, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter,  
Nature 402, 761-768, 1999  
A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
A:Reference number: A84420; MUID:20083487; PMID:10617197  
A:Accession: H84451  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <STO>  
A:Cross-references: GB:AB002093; NID:G4406766; PIDN:AAD20077.1; GSPDB:GN00139  
C:Genetics:  
A:Gene: At2G03750  
A:Map position: 2  
C:Superfamily: alcohol sulfotransferase

Query Match 26.5%; Score 545; DB 2; Length 331;  
Best Local Similarity 38.4%; Pred. No. 1.5e-31;  
Matches 118; Conservative 61; Mismatches 118; Indels 10; Gaps 8;

QY 36 EDSKMYKRYREIVSSLPNDYWGDTMRLYKGFQMGYLVPGIMAFENFKAETDILIT 94  
DB 15 KDNVSVQETKNLITSLPSKDFMGYGLNYKGCWYYPNTLQAVLDVQKHFKPRDITILA 74  
QY 95 TLPKAGTTWKALTFAILTRDVNHPSSPHLLFPNPHSCVNLEYLYMGRENTMPDLD 153  
DB 75 SLPKGTTWLKSLIFAVVHREKYRTPQTHPILLQNPDLVPLFVELVANSQ--IPDLA 132





```

11 REEEKPSEETKILISSUPWEIDYLGKLFNYEGWYSEDIQSPINHTGFOQETDII 70
93 LTTLPKAGTTWKALTFAILLTRDVNHPSSPHLLFFNPHSCVONLEY-LYMGRENTMPD 151
71 LASFYKSGTTWKALTFAIVORSKHSLEDHQHLLHNPHEIVPNLELDLYL--KSKRPD 128
152 LDML---NESPRLPAGHIPYSLPASPVLKSGTKIINISNRKSTFVSFWKF--GNLNP 205
129 LTKFLSSSSSPRLSTEMSLDPLQVPLKNCILXIVYCRNVKDVWVSVYFRQSKITR 188
206 DKLLDLEKSVDFIAGISFCGPENNFQAEFTNAASTN-SNLLLSYEEMLEKPVENVKL 264
189 AEDYSLEAIFPFCNGVTLHGPFWDHALSYWRGSLDEPKHFLFMYEDLKAEPQTQVRL 248
265 AEFMCGGTTDDEKOGIVDEIVKLCSPDLNKNQVKNKNGSSYNKIDNKHFFRKGEVRDW 324
249 AEFLOCPPTKEEDSGVDKILELCSLNSLSVEINKTETS--SRVDFKSYFRKGQVGDW 306
325 ANYLTSEMIMKLE 337
307 KSYMTPENVYDKID 319

```

## RESULT 14

```

A40216
flavonol 4'-sulfotransferase - Flaveria chloraefolia
C:Species: Flaveria chloraefolia
C>Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 20-Sep-1999
C:Accession: A40216
R:Varin, L.; Deluca, V.; Ibrahim, R.K.; Brisson, N.
Proc. Natl. Acad. Sci. U.S.A. 89, 1286-1290, 1992.
A:Title: Molecular characterization of two plant flavonol sulfotransferases.
A:Reference number: A40216; MUID:92159034; PMID:1741382
A:Accession: A40216
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-320 <VAR>
A:Experimental source: terminal bud
A>Note: sequence inconsistent with the nucleotide translation
A>Note: sequence extracted from NCBI backbone (NCBIN:82216, NCBIPI:82217)
C:Superfamily: alcohol sulfotransferase

```

```

Query Match 24.8%; Score 510.5; DB 2; Length 320;
Best Local Similarity 37.6%; Pred. No. 4.3e-29;
Matches 120; Conservative 58; Mismatches 110; Indels 31; Gaps 9;

```

```

QY 35 EDSRMVKRYREIVSSLP---SNDYWGDTMLYKGFQWQGYLVPGIMAFEDNFKARETD 90
DB 2 ETTKTOPESMAEMIKKLPQHTCSSLKGRITLYKYQDFWGLQNNIEGAILAQOSFKAPDD 61
QY 91 IILTLPKAGTTWKALTFAILLTRDVNHPSSPHLLFFNPHSCVONLEYLYMGRENTMP 150
DB 62 VFLCSVPKSGTTWLKALAYAVTRE--KDEFTSPLLTWIPENCIPYIE-----K 109
QY 151 LDML--NESPRLF---AGHIPYSLPASPVLKSGTKIINISNRKSTFVSFWKFGNLI 205
DB 110 DLAKIVDNQNNSCFTPMATHYVLPKSLALNCKWYIYRNKIDVIVSYFHFGEITK 169
QY 206 DKLLD--LEKSVDFIAGISFCGPENNFQAEFTNAASTNLL--LSYEEMLEKPVENVK 262
DB 170 LPLEADAPFEAFDEFYHGISQFGPYWDHLLGYKWSLERPEVILFLKYEDVKDPTSNV 229
QY 263 KLAFFMCGGTTDDEKOGIVDEIVKLCSPDLNKNQVKNKNGSSYNK----IDNKHFFRK 318
DB 230 RLAEFIGYPTFEEKEGVIESIKLCSFENLNLNLEVNKSG---NSKGFLPIENRLYFRK 286
QY 319 GEVRDWNALYLTSEMIMKLE 337
DB 287 AKOGDWKNYFTDENTEKID 305

```

## RESULT 15

```

T07832
probable steroid sulfotransferase (EC 2.8.2.15) 2 - rape
C:Species: Brassica napus (rape)
C>Date: 14-May-1999 #sequence_revision 14-May-1999 #text_change 08-Oct-1999
C:Accession: T07832
R:Richard, M.; Nicolle, L.; Varin, L.
submitted to the EMBL Data Library, April 1997
A:Reference number: Z16161
A:Accession: T07832
A>Status: translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-324 <RIC>
A:Cross-references: EMBL:AF000306; NID:g3420005; PIDN:AAC63112.1; PID:g3420006
C:Superfamily: alcohol sulfotransferase
C:Keywords: sulfotransferase

```

```

Query Match 24.3%; Score 500.5; DB 2; Length 324;
Best Local Similarity 38.0%; Pred. No. 2.3e-28;
Matches 117; Conservative 58; Mismatches 114; Indels 19; Gaps 10;

```

```

QY 37 DSKMYKRYREIVSSLPNDYW-GDTMLYKGFQWQGYLVPGIMAFEDNFKARETDIILT 95
DB 13 DENLTOKTKDLISLSPSEKGLVQCMYQQRWHTQALLOGLTCQKHFEAKDSIILVT 72
QY 96 LPKAGTTWKALTFAILLTRDVNHPSSPHLLFFNPHSCVQN--LEYLYMGRE-NTMPDLD 153
DB 73 NPKSGTITWLKALVFAL----INRHKFPVYSVAIL---SCYQSALLVPFLGRSILRSPDFD 125
QY 154 MLN-ESPRLPAGHIPYSLPASPVLKSGTKIINISNRKSTFVSFWKFGNLIINPDKLLD-- 210
DB 126 FSQLSLSPRLMNTTHISLSPSPSVKSSCKIVYCCRNPKDMFVSLMHFGKGLAPEETADYP 185
QY 211 LEKSVDFIAGISFCGPENNFQAEFTNAASTNSN-LLLLSYEEMLEKPVENVKKLAFFMG 269
DB 186 IEKAVEAFCCGKFIQGFWDHVLVYVWASLENKVLVSYBEPKKTGTETIKRIAEFLG 245
QY 270 CGFTDDEKOGIVDEIVKLCSPDLNKNQVKNKNGSSYNKIDNKHFFRKGEVRDWNALY 329
DB 246 CGLVGEE---VRAIVKLCSPESLSLEVNREG-KLPSCMETRAPFRKGEVGGWRDILT 300
QY 330 SEMIKKLE 337
DB 301 ESIAEVID 308

```

```

Search completed: March 20, 2004, 01:39:46
Job time : 17 secs

```

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:35:33 ; Search time 11 Seconds  
(without alignments)  
1850.857 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059

Sequence: 1 TRGITGVACPLIMAGIL.....CILSYFVINPILKKKK 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	537.5	26.1	326	1	FSTL_ARATH
2	510.5	24.8	320	1	P4STL_FLACH
3	491.5	23.9	312	1	P3ST_FLABI
4	473	23.0	309	1	FSTL_FLABI
5	455	22.1	311	1	F3ST_FLACH
6	309	15.0	295	1	SUOE_BOVIN
7	304	14.8	302	1	S1C2_HUMAN
8	255.5	12.4	304	1	SUAC_RAT
9	255	12.3	294	1	STOE_HUMAN
10	253	12.3	295	1	SUO3_RAT
11	252	12.2	295	1	SUO1_RAT
12	252	12.2	295	1	SUO2_RAT
13	250	12.1	295	1	SUO6_RAT
14	248.5	12.1	291	1	SUAR_MOUSE
15	248.5	12.1	295	1	SUOT_MOUSE
16	245	11.9	291	1	SUAR_RAT
17	243.5	11.8	284	1	SUHA_HUMAN
18	239.5	11.6	296	1	SUOE_CAVPO
19	238	11.6	296	1	STK2_RAT
20	233	11.3	296	1	STK1_RAT
21	232	11.3	299	1	SUDY_RAT
22	229	11.1	294	1	SUHA_MACFA
23	228.5	11.1	294	1	SUPP_BOVIN
24	228	11.1	296	1	S1C1_RABIT
25	225.5	11.0	295	1	SUF2_HUMAN
26	224	10.9	296	1	S1C1_HUMAN
27	223.5	10.8	295	1	SUF1_HUMAN
28	222.5	10.8	295	1	SUPM_HUMAN
29	221	10.7	295	1	SUPP_MACFA
30	213	10.3	285	1	SUH2_MOUSE
31	212	10.3	285	1	SUH2_RAT
32	211.5	10.3	283	1	SUHS_RAT
33	211.5	10.3	286	1	SUHB_CAVPO

34	208.5	10.1	286	1	SUHA_CAVPO
35	205.5	10.0	283	1	SUHA_RAT
36	205.5	10.0	285	1	SUH3_RAT
37	201	9.8	285	1	SUH1_MOUSE
38	187	9.1	284	1	S4A1_HUMAN
39	182	8.8	284	1	S4A1_MOUSE
40	119	5.8	1628	1	YATE_SCHPO
41	119	5.8	1726	1	MSPI_PLAFC
42	119	5.8	1726	1	MSPI_PLAPP
43	118.5	5.8	944	1	UVRA_MYCFU
44	115.5	5.6	1701	1	MSPI_PLAPP
45	115.5	5.6	1701	1	MSPI_PLAFM

#### ALIGNMENTS

#### RESULT 1

ID	FSTL_ARATH	STANDARD	PRT	326 AA
AC	P52839; O9SJW2;			
DT	01-OCT-1996 (Rel. 34, Created)			
DT	28-FEB-2003 (Rel. 41, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Flavonol sulfoltransferase-like (EC 2.8.2.-) (RaRO47).			
GN	AT2G3760 OR F19B11.21.			
OS	Arabidopsis thaliana (Mouse-ear cress).			
OC	Eukaryota; Viridiplantae; Streptophyta;			
OC	Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;			
OC	eurosid1 II; Brassicales; Brassicaceae; Arabidopsis.			
OX	NCBI_TaxID=3702;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV, Columbia;			
RX	MEDLINE=96270377; PubMed=8639757;			
RA	Lacomme C., Roby D.;			
RT	"Molecular cloning of a sulfoltransferase in Arabidopsis thaliana and			
RT	regulation during development and in response to infection with			
RT	pathogenic bacteria.";			
RL	Plant Mol. Biol. 30:995-1008(1996).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV, Columbia;			
RX	MEDLINE=20083487; PubMed=10617197;			
RA	Lin X., Kaul S., Rounsley S.D., Shea T.P., Benito M.-I., Town C.D.,			
RA	Fujii C.Y., Mason T.M., Bowman C.L., Barnstead M.E., Feldblum T.V.,			
RA	Buell C.R., Ketchum K.A., Lee J.J., Ronning C.M., Koo H.L.,			
RA	Moffat K.S., Cronin L.A., Shen M., Pai G., Van Aken S., Umayani L.,			
RA	Tallon L.J., Gill J.B., Adams M.D., Carrera A.J., Creasy T.H.,			
RA	Goodman H.M., Sonerville C.R., Copenhaver G.P., Preuss D.,			
RA	Nierman W.C., White O., Eisen J.A., Salzberg S.L., Fraser C.M.,			
RA	Venter J.C.;			
RA	"Sequence and analysis of chromosome 2 of the plant Arabidopsis			
RA	thaliana.";			
RL	Nature 402:761-768(1999).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CV, Columbia;			
RX	MEDLINE=22954850; PubMed=14593172;			
RA	Yamada K., Lim J., Dale J.M., Chen H., Shinn P., Palm C.J., Cheuk R.F.,			
RA	Southwick A.M., Wu H.C., Kim C.J., Nguyen M., Pham P.K., Cheuk R.F.,			
RA	Karlin-Newmann G., Liu S.X., Lam B., Sakano H., Wu T., Yu G.,			
RA	Miranda M., Quach H.L., Tripp M., Chang C.H., Lee J.M., Toriumi M.J.,			
RA	Chen M., Tang C.C., Onodera C.S., Deng J.M., Akiyama K., Carninci P.,			
RA	Arakawa T., Ban H., Banno F., Bowser L., Brooks S.Y., Carninci P.,			
RA	Chao Q., Choy N., Enju A., Goldsmith A.D., Gurjal M., Hansen N.F.,			
RA	Hayashizaki Y., Johnson-Hopson C., Huan V.W., Iida K., Karnes M.,			
RA	Khan S., Koeseema B., Ishida J., Jiang P.X., Jones T., Kawai J.,			
RA	Kamiya A., Meyers C., Nakajima M., Narusaka M., Seki M., Sakurai T.,			
RA	Satou M., Tamse R., Vaysberg M., Wallender E.K., Wong C., Yamamura Y.,			
RA	Yuan S., Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;			
RT	"Empirical analysis of transcriptional activity in the Arabidopsis			
RT	genome.";			

P50234 cavia porce  
P22789 rattus norv  
P50235 mus musculu  
Q9br01 homo sapien  
Q9dc97 mus musculu  
Q09779 schizosacch  
P04934 plasmodium  
P50495 plasmodium  
Q98p12 mycoplasma  
P3819 plasmodium  
P08569 plasmodium







```

DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Flavonol 3-sulfotransferase (EC 2.8.2.-) (F3-ST).
OS Flaveria bidentis.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;
OC campanulids; Asterales; Asteraceae; Asteroideae; Tageteae; Flaveria.
OX NCBI_TaxID=4224;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95083751; PubMed=7991681;
RA Ananvoranich S., Varin L., Gulick P., Ibrahim R.;
RT "Cloning and regulation of flavonol 3-sulfotransferase in cell-
RT suspension cultures of Flaveria bidentis.";
RL Plant Physiol. 106:485-491(1994).
CC -!- FUNCTION: Catalyzes the sulfate conjugation of quercetin,
CC rhamnetin and isochanetin but not kaempferol. O-sulfation of
CC position 3 of flavonol. May play a role in auxin transport.
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Highest in shoot tips and lowest in mature
CC leaves and roots.
CC -!- SIMILARITY: Belongs to the plant sulfotransferase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U10275; AAA61638.1; -.
DR HSP; P50224; 1CJM.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
DR Transferrase. 138 153 PAPS-BINDING SITE (BY SIMILARITY).
KW BINDING 312 AA; 36457 MW; 29D174E8CB663FC2 CRC64;
FT SEQUENCE
SQ
Query Match 23.9%; Score 491.5; DB 1; Length 312;
Best Local Similarity 38.2%; Pred. No. 7.4e-28;
Matches 118; Conservative 46; Mismatches 114; Indels 31; Gaps 9;
QY 46 EIVSSLPND--DYGDWTWRYK--GFWQMGVLPVGVAFEDNFKARETDIILTLPKAGT 101
DB 3 DIKTLPGHTCSFLKQFTLYKYQDVVNHQFLEGRMLSEQTAKHPNDVFLASYPKSGT 62
QY 102 TWTKALFAITLROVNHPSPTPLFFNPCHSCVQNLVYMGRENTMPDMLNESPRL 160
DB 63 TWKALAFATITREKFDSD--TSPLLTTPHDCIPLLE-----KDEKIQENQRN 110
QY 161 ----LFAGHIPYSLLPASVLKSGTKIINISNRKSTFVSFWKFGNLINPKL----LDLE 212
DB 111 SLYTPISITFHYKSLPESARTSNCKIIVYIYRNKMDIVSYTHP--LRQIVKLSVEEAPFE 168
QY 213 KSVDFEASGISFCGPEWFOAEFTNAASTNSL--LLLSYEEMLEKPVENVKLAEFMGCG 271
DB 169 EAVDEFCQGISCGPYWEHILGYKASLEKPEIFFLFKYEDMKDQVPVSVKKLADIGHP 228
QY 272 FTDDEKQGIYDEIVKLCSPDNLKQVKNKNGSSYNSK---INKHFRKGEVRDWNLYL 328
DB 229 FTPKEEAGVIENTIKLCSPEKLSLEVNSGMRHPEAHSENRLYFRKGDGDKWQNYF 288
QY 329 TSEMICKLE 337
DB 289 TDEMIKID 297
RESULT 4
FSTL FLABI
ID FSTL FLABI STANDARD; PRT; 309 AA.
DT 01-OCT-1996 (Rel. 34, Created)
FSTL FLACH
ID FSTL FLACH STANDARD; PRT; 311 AA.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Flavonol 3-sulfotransferase (EC 2.8.2.-) (F3-ST).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OS Flaveria chloraefolia.
OX NCBI_TaxID=4224;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95232186; PubMed=7716232;
RA Ananvoranich S., Gulick P., Ibrahim R.K.;
RT "Flavonol 3-sulfotransferase-like cDNA clone from Flaveria bidentis.";
RL Plant Physiol. 107:1019-1020(1995).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC -!- SIMILARITY: Belongs to the plant sulfotransferase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; U10277; AAA87399.1; -.
DR HSP; P50224; 1CJM.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransfer; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
DR Transferrase. 138 153 PAPS-BINDING SITE (BY SIMILARITY).
KW BINDING 309 AA; 35917 MW; 598A9C9EBFB75303 CRC64;
FT SEQUENCE
SQ
Query Match 23.0%; Score 473; DB 1; Length 309;
Best Local Similarity 36.4%; Pred. No. 1.5e-26;
Matches 111; Conservative 59; Mismatches 109; Indels 26; Gaps 9;
QY 46 EIVSSLPND--YW---GDTMELYKGFQMGVLPVGVAFEDNFKARETDIILTLPKAGT 101
DB 3 EILKTLPGHTCSWLKHKIIMYKYQDFWTSKOLLEGLTMAQQSFKAEPDVFCLCAPKGT 62
QY 102 TWTKALFAITLROVNHPSPTPLFFNPCHSCV---QNLVYMGRENTMPDMLNLE 157
DB 63 TWKALAFATITRE--NFDES--TSPLLKLVHECVFPLERQVEEIEHNRES----- 111
QY 158 SPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFVSFWKFGNLINPKL----LDLEK 213
DB 112 SLPLVATHLPVSLPESVIASNCRMVYIYRNKDVIVSNYHF--LREAFKLSMEDAPFEE 169
QY 214 SVDIFASGISFCGPEWFOAEFTNAA--STNSNLLLSYEEMLEKPVENVKLAEFMGCGF 272
DB 170 TFEDEFNGNSSVGPYWDHILGYKASLMDPKILFLKTEDLKSEPISNVKLAEFIGYFF 229
QY 273 TDDEEKQGIYDEIVKLCSPDNLKQVKNKNGSSYNSKIDNKHFFRKGEVRDWNLYLSEM 332
DB 230 SNDEKAGVIENTINNCSFENLSLEVNTKRPKGMLENRLYKRGDGDGDKWQNYFTNEM 289
QY 333 IKKLE 337
DB 290 KEKID 294
RESULT 5
FSTL FLACH
ID FSTL FLACH STANDARD; PRT; 311 AA.
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Flavonol 3-sulfotransferase (EC 2.8.2.-) (F3-ST).
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

```



```

or send an email to license@sib.ch).
CC EMBL; M54942; AAA30679.1; -
CC DR EMBL; X56395; CAA39806.1; -
CC DR PIR; S29045; S29045.
CC DR HSP; P49891; IAQU.
CC DR InterPro; IP000863; Sulfotransferase.
CC DR Pfam; PF00685; Sulfotransferase; 1.
CC DR ProDom; PD001218; Sulfotransferase; 1.
CC KW Transferase; Phosphorylation; Steroid-binding.
CC FT MOD_RES 216 216 PHOSPHORYLATION (BY PKA) (POTENTIAL).
CC FT MOD_RES 228 228 PHOSPHORYLATION (BY PKA) (POTENTIAL).
CC FT CONFLICT 117 117 S -> E (IN REF. 2).
CC FT CONFLICT 171 171 S -> E (IN REF. 2).
CC FT CONFLICT 248 248 M -> Q (IN REF. 2).
CC SQ SEQUENCE 295 AA; 34640 MW; 69147C73146913FD CRC64;

Query Match 15.0%; Score 309; DB 1; Length 295;
Best Local Similarity 31.0%; Pred. No. 6.3e-15;
Matches 95; Conservative 50; Mismatches 113; Indels 48; Gaps 14;

QY 49 SSIPLS-NDYKGDY--NRLYKGFQWQGYLYPGIMAFEDNFKARETDIILTLPLKAGTTWK 105
DB 3 SSIPLSFDYFKLGGIPMTKIEGFHN-----EFKARPDVLVIVTPKSGTTWLS 55
QY 106 ALTFALTR-DVNHPSPTPLLPFNPHSCVQNLLEYLYMGRENTMPDLDMLNE--SPRLF 162
DB 56 EILCMYNGDVEKCKED---VIFN-----RVPLECSTEHVMKGVKQLNEMASPRIV 105
QY 163 AGHIPYSLLPASVLRSGTKIINISNRKSTFVSFWKFGNLI-----NPKLLD-LEKVD- 216
DB 106 KSHLPVKLLPVGFWEKKNKIIYLSRNKADVVVSYFLIMVTAIPDPDSFODFVEKFMOG 165
QY 217 --IFASGISFCGPEWFOAFTNAASTNSNLLLSYEEMLEKPVENVKLAEPFGCGFTD 274
DB 166 EYVYGSWFETHKSWNE-----KSNQPVLFLEVEDKMKIRKVMKLEFLGRASD 217
QY 275 DEEQGIIVDEIVKLSFDMKLNQVKNQNGSSYNKIDNKH---FPRKEVDRWANYLTSE 331
DB 218 E-----LVDKIKHTSFQEMKNKP-STNYTLTLDVDMVKNQVSPFMRKGDVGDMKNHETVA 271
QY 332 MINKLE 337
DB 272 LNEKFD 277

RESULT 7
SIC2_HUMAN STANDARD; PRT; 302 AA.
AC O75837;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Sulfotransferase 1C2 (EC 2.8.2.-) (SULT1C) (SULT1C#2).
GN SULT1C2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Fetal lung;
RX MEDLINE=99069375; PubMed=9852044;
RA Sakakibara Y., Yanagisawa K., Katafuchi J., Ringer D.P., Takami Y.,
RA Nakayama T., Suiko M., Liu M.-C.;
RT "Molecular cloning, expression, and characterization of novel human
RT SULT1C sulfotransferases that catalyze the sulfonation of
RT N-hydroxy-2-acetylaminofluorene."
RL J. Biol. Chem. 273:33929-33935(1998).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=20247255; PubMed=10783263;
RA Freimuth R.R., Raftogiannis R.B., Wood T.C., Moon E., Kim U.-J., Xu J.,

```

```

RA Siciliano M.J., Weinshilboum R.M.;
RT "Human sulfotransferases SULT1C1 and SULT1C2: cDNA characterization,
RT gene cloning, and chromosomal localization.";
RL Genomics 65:157-165(2000).
CC -!- FUNCTION: Catalyzes the sulfate conjugation of many drugs,
CC xenobiotic compounds, hormones, and neurotransmitters. May be
CC involved in the activation of carcinogenic hydroxylamines. Shows
CC activity towards p-nitrophenol and N-hydroxy-2-acetylami-
CC fluorene (N-OH-2AAF).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- TISSUE SPECIFICITY: Expressed at high levels in fetal lung and
CC kidney and at low levels in fetal heart, adult kidney, ovary and
CC spinal chord.
CC -!- SIMILARITY: Belongs to the sulfotransferase family.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).

CC EMBL; AF055584; AAC95519.1; -
CC DR EMBL; AF186263; AAF72810.1; -
CC DR HSSP; P50224; 1CJM.
CC Genew; HGNC:11457; SULT1C2.
CC DR GO; GO:0008146; P:sulfotransferase activity; NAS.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransferase; 1.
CC ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
FT CONFLICT 5
FT SEQUENCE 302 AA; 35534 MW; DD2F4865932E3B56D CRC64;

Query Match 14.8%; Score 304; DB 1; Length 302;
Best Local Similarity 30.0%; Pred. No. 1.5e-14;
Matches 93; Conservative 49; Mismatches 114; Indels 54; Gaps 13;

QY 59 DTMRLYKGFQWQGYLYPGIMAFEDNFKARETDIILTLPLKAGTTWKALTFALTR-DVN 117
DB 30 DTCDIWDKIW-----NFOAKPDDLIISTYPRAGTTWTQEIIVELIQNEGJVE 75
QY 118 HPS-SPTH---PLLFPNPHSCVQNLLEYLYMGRENTMPDLDMLNESPLFAGHIPYSLLPA 173
DB 76 KSKRAPTHQRPFFLEMKIPLSGGLE-----QAEAMP-----SPRLKTHLPLHLLPP 123
QY 174 SVLKSGTKIINISNRKSTFVSFWKFGNLIINPKLLDL---EKSVDFASGISFCGPEW 229
DB 124 SLLERNCKIIYARNPKDNVSYTHFORM---NKAIPAPGTWEEYFETFLAG-KVCWGSW 179
QY 230 NFOASFTNAASTNSNLLLSYEEMLEKPVENVKLAEPFGCGFTDDEKQGIIVDEIVKLC 289
DB 180 HEHVKGWWEAKDKRILYLFYEDMKQPKKHEIOKLAEFIQKLLDK-----VLDKIVHYT 234
QY 290 SPDNLNQVKNQNGSSYNKIDN---KHFFRKGEVDRWANYLTSEMIKKLETAGKINESE 346
DB 235 SFDVWK-QNPWANYSSIPAEIMDHSISPFWRKGAVGDKKKHFTVAQNERFD-----EDY 287
QY 347 KHLISNKNLT 356
DB 288 KKKMTDTRLT 297

RESULT 8
SUAC RAT STANDARD; PRT; 304 AA.
AC SUAC RAT
ID SUAC237;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE N-hydroxyaryamine sulfotransferase (EC 2.8.2.-) (HAST-I).
GN ST1C1.

```



"Estrogen sulfotransferase of the rat liver: complementary DNA cloning and age- and sex-specific regulation of messenger RNA."; Mol. Endocrinol. 6:589-597(1992).

(2)

SEQUENCE FROM N.A.

RP TISSUE=Liver;

RC MEDLINE=95161123; PubMed=7857871;

RX Palany J.L., Krasnykh V., Mikhoeva G., Palany C.N.;

RA "Isolation and expression of an isoform of rat estrogen sulfotransferase.";

RT J. Steroid Biochem. Mol. Biol. 52:35-44(1995).

RL -!- FUNCTION: Sulfation of estrone and estradiol. May control the level of the estrogen receptor by sulfonylating free estradiol.

CC -!- CATALYTIC ACTIVITY: 3'-phosphadenylylsulfate + estrone = adenosine 3',5'-bisphosphate + estrone 3-sulfate.

CC -!- SUBUNIT: Homodimer (by similarity).

CC -!- SUBCELLULAR LOCATION: Cytoplasmic.

CC -!- TISSUE SPECIFICITY: LIVER OF YOUNG MATURE MALES AND UTERUS.

CC -!- DEVELOPMENTAL STAGE: Expressed only in the liver of young adult animals (100 days old) and is absent in the prepubertal male (27 days old), senescent male (800 days old) and female liver.

CC -!- INDUCTION: Induced by androgens and suppressed by estrogens. The expression is under the influence of pituitary growth hormone and thyroid hormone. Is regulated by progesterone in the uterus.

CC -!- SIMILARITY: Belongs to the sulfotransferase family.

CC

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

EMBL; M86758; AAA41128.1; -

DR EMBL; S76489; AAB33441.1; -

DR PIR; A41930; A41930.

DR HSP; P49891; IAQU.

DR InterPro; IPR000863; Sulfotransferase.

DR Pfam; PF00685; Sulfotransfer; 1.

DR ProDom; PD001218; Sulfotransferase; 1.

DR Transfaser; Steroid-binding; Multigene family.

FT BINDING 259 265 PAPS BINDING SITE (POTENTIAL).

FT CONFLICT 150 150 P -> Q (IN REF. 2).

FT CONFLICT 238 238 T -> I (IN REF. 2).

FT CONFLICT 295 295 L -> P (IN REF. 2).

FT SEQUENCE 295 AA; 35415 MW; A77807A21DD2E7EB CRC64;

QY

Query Match 12.3%; Score 253; DB 1; Length 295;

Best Local Similarity 26.7%; Pred. No. 5.9e-11;

Matches 85; Conservative 41; Mismatches 112; Indels 80; Gaps 12;

QY 40 MYKRYEIVGSLPSNDYWGTMELYGFWQGVLPVGMAFEDNFKARETDILTLPKA 99

DB 20 MDKRETK-----YWEDEV-----ETFLAPDDLLIVTPKS 49

QY 100 GTTWTKALTEAILTR-DVNHPSPTPLLFNPHSCVQNLEYLYMGRENTMPDL-----D 153

DB 50 GSTWISIVDMVYKEGDV-----EKCKEDALF-----NRIPDLECRNED 88

QY 154 MLN-----ESPLRFAGHIPSLPASVLKSGTKLINISRNKSTFVSPWKEGNLI-- 203

DB 89 LINGIKOLKEKESPRIKTHLPKALLPASFEWKCKVIYLCRNAKDVVVVYIYFLIMKS 148

QY 204 --NPKLLDLEKSDVIDFASGISFCGPBNWFOAEFTNAASTNSNLLLSYBEMLEKPVENV 261

DB 149 YPNPKSSEP---VEKFWEGQVPVG-SWYDHVKSWEKSKNSRVLFWFYEDMKEDIRREV 204

QY 262 KKLAEFMGCGFTDDEEKQGVDEIVKLCSPDNLKQKQNKNGSSYNSKIDNK--HFRKKG 319

DB 205 VKLIETFL-----ERDPSSAELVDRIIOHTSFQEMKNNPCTNYSMLPMTIDLVKVSFPMRKG 259

QY 320 EVRDWNVYLTSMTIKKLE 337

```
Db 260 IVGDWKNHFFPALRERPE 277
RESULT 11
SUO1_RAT
ID SUO1_RAT STANDARD; PRT; 295 AA.
AC PS2844;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Estrogen sulfotransferase, isoform 1 (EC 2.8.2.4) (EST-1)
DE (Sulfotransferase, estrogen-prefering) (Estrone sulfotransferase).
GN STEL OR SULT1S1
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Fischer 344; TISSUE=Liver;
RX MEDLINE=96305357; PubMed=6688469;
RA Rikke B.A.; Roy A.K.;
RT "Structural relationships among members of the mammalian
sulfotransferase gene family."
RL Biochim. Biophys. Acta 1307:331-338(1996).
CC -!- FUNCTION: Sulfation of estrone and estradiol. May control the
level of the estrogen receptor by sulfurylating free estradiol (By
similarity).
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + estrone =
adenosine 3',5'-bisphosphate + estrone 3-sulfate.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the sulfotransferase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U50204; AAB07680.1; -
CC HSSP; P49891; 1AQU.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransfer; 1.
CC ProDom; PD001218; Sulfotransferase; 1.
CC Transferrase; Steroid-binding.
FT BINDING 259 265 PAPS BINDING SITE (POTENTIAL).
SQ SEQUENCE 295 AA; 35509 MW; 696A12FDA92A12E CRC64;

Query Match 12.2%; Score 252; DB 1; Length 295;
Best Local Similarity 26.7%; Pred. No. 7e-11;
Matches 85; Conservative 41; Mismatches 112; Indels 80; Gaps 12;

QY 40 MYKREIVSLPSNDYGDITMLYKGFQWQYLVPGIMAFEDNFKARETDILTLPKA 99
Db 20 MDKRRTK-----YWDI-----ETFLARPDDLIVTPKS 49
QY 100 GTTWTKALTFAILTR-DVNHPSSTPHLLFFNPHSCVQNLVYMGRENTMPDL-----D 153
Db 50 GSTWISSEIVDMYKGDV-----EKCKEDALF-----NRIPDLECRNED 88
QY 154 MLN-----ESPRLPAGHYPYLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI-- 203
Db 89 LINGIKQKEREKSPRIVTHLPKLLPASFWKKNCKIILYCRNAKDVVVSYFFFLIMKS 148
QY 204 --NPKLLDLEKSVDFIFAGSIFGCPGPNWFOAFTNAASTNSNLLLSYEMLKPVNY 251
Db 149 YENPKSFSEF---VEKMEGQVPYG-SWYDHVKSWWEKSKGSRVLFMFYEDMKEDJRRVY 204
QY 262 KKLAEFMCGGFTDDSEKQIVDEIVKLGCSFDNLKNQKNNGSYNSKIDNK--HFRK 319
```

```
Db 205 VKLIEFL-----ERDFSAELVDRIIQTSTFQEMKNPNCTNYSMLPETMIDLVSPFMRKG 259
QY 320 EYRDWANYLTSEMINKLE 337
Db 260 IVGDWKNHFFPALRERPE 277

RESULT 12
SUO2_RAT
ID SUO2_RAT STANDARD; PRT; 295 AA.
AC PS2845;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE Estrogen sulfotransferase, isoform 2 (EC 2.8.2.4) (EST-2)
DE (Sulfotransferase, estrogen-prefering) (Estrone sulfotransferase).
GN STE2.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=Fischer 344; TISSUE=Liver;
RX MEDLINE=96305357; PubMed=6688469;
RA Rikke B.A.; Roy A.K.;
RT "Structural relationships among members of the mammalian
sulfotransferase gene family."
RL Biochim. Biophys. Acta 1307:331-338(1996).
CC -!- FUNCTION: Sulfation of estrone and estradiol. May control the
level of the estrogen receptor by sulfurylating free estradiol (By
similarity).
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + estrone =
adenosine 3',5'-bisphosphate + estrone 3-sulfate.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the sulfotransferase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; U50205; AAB07681.1; -
CC HSSP; P49891; 1AQU.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransfer; 1.
CC ProDom; PD001218; Sulfotransferase; 1.
CC Transferrase; Steroid-binding.
FT BINDING 259 265 PAPS BINDING SITE (POTENTIAL).
SQ SEQUENCE 295 AA; 35364 MW; 149B5C9D46039AAF CRC64;

Query Match 12.2%; Score 252; DB 1; Length 295;
Best Local Similarity 26.0%; Pred. No. 7e-11;
Matches 85; Conservative 44; Mismatches 131; Indels 44; Gaps 12;

QY 49 SSLES-----NDYWGDTM-RLYKGFQWQYLVPGIMAFEDNFKARETDILTLPKAGT 102
Db 3 TSMPEYVEVGFDFHGLMDKLFYKWE-----DVETFSARPDLLVVVTPKSGST 52
QY 103 WTKALTFAILTR-DVNHPSSTPHLLFFNPHSCVQNLVYMGRENTMPDL--NESP 159
Db 53 WIGSEIVDMYKGDVEKCKEDA-----IFN-----RIPYLECRNEDLINGIKQLEKESF 102
QY 160 RLFAGHYPYLLPASVLKSGTKIINISRNKSTFVSFWKFGNLI-----NPKLLDLEKSV 215
Db 103 RIVKTHLPKLLPASFWKKNCKIILYCRNAKDVVVSYFFFLIMKSYNPKSFSEF---V 159
QY 216 DIFASGIFSGCPGPNWFOAFTNAASTNSNLLLSYEMLKPVNYVKLAEFMCGGFTDD 275
```



```
Db 160 EKFMEGQVPYG-SWYDHVKSWEKSKSRVLPFYEDMKEDIRREVVKLIEFL-----BR 213
Qy 276 BEKQIGVDEIVKLSFNLKQKQVKNKSSYNKIDNK--HFFRKGEVRDWNLYTSEMI 333
Db 214 DPAELVDRIIQTSTFOEMKKNPCTNYSMLPETMIDLKVSFPMRKGIVGDWRNHPPEARL 273
Qy 334 KLE 337
Db 274 ERPE 277

RESULT 13
SU06_RAT
ID SU06_RAT STANDARD; PRT; 295 AA.
AC P49890;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Estrogen sulfotransferase, isoform 6 (EC 2.8.2.4) (EST-6)
DE (Sulfotransferase, estrogen-prefering) (Estrogen sulfotransferase).
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
CX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=95161323; PubMed=7857871;
RA Falany J.L., Kraenkyh V., Mikheeva G., Falany C.N.;
RT "Isolation and expression of an isoform of rat estrogen
sulfotransferase.";
RL J. Steroid Biochem. Mol. Biol. 52:35-44 (1995).
CC -!- FUNCTION: Sulfation of estrone and estradiol. May control the
level of the estrogen receptor by sulfurylating free estradiol.
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + estrone =
adenosine 3',5'-bisphosphate + estrone 3-sulfate.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC -!- DEVELOPMENTAL STAGE: Expressed in males rats, but not in females.
CC -!- INDUCTION: Induced by androgens and suppressed by estrogens.
CC Expression is under the influence of pituitary growth hormone and
thyroid hormone.
CC -!- SIMILARITY: Belongs to the sulfotransferase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
CC
CC EMBL; S76490; A833442.1; -
CC HSP; P49891; IAOU.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransfer; 1.
CC ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid-binding; Multigene family.
FT BINDING 259 AA; 35302 MW; 9FBD5861AFDC9B05 CRC64;
SQ SEQUENCE 295 AA; 35302 MW; 9FBD5861AFDC9B05 CRC64;

Query Match 12.1%; Score 250; DB 1; Length 295;
Best Local Similarity 28.0%; Pred. No. 9.7e-11;
Matches 85; Conservative 45; Mismatches 130; Indels 44; Gaps 12;

Qy 49 SSIPS-----NDYMGDTW-RLYKGFQMGVLPVGMAFEDNFKARETDIILTLPKAGTT 102
Db 3 TSMPEYEVGDFHGVLDKLFKWE-----DVEFSARPDDLVLVTPKSGST 52
Qy 103 WTKALTFAILTR-DVNHPSSTPHLLFFNPHSCVQNLEYLYMGRENTWPDLDML--NESP 159
Db 53 WIGEIVDMYIKSGDVEKCKEDA----IFN-----RIPYLECRNEDLINGIKLKEKESP 102
```

```
Qy 160 RLFAGHIPYSLLPASVLKSGTKIINISRNKRKSTFVSFWKFNLI-----NPKLLDLEKSV 215
Db 103 RIVKTHLPKALLPASFWKEKCKIIVLCENAKDVVVSYVFFLIKSPNPKSFSEF---V 159
Qy 216 DIPASGISFCGPENFQAETNAASTNSNLLLSVEEMLEKPVENVKLAEFMCGFTDD 275
Db 160 EKFMEGQVPYG-SWYDHVKSWEKSKSRVLPFYEDMKEDIRREVVKLIEFL-----ER 213
Qy 276 BEKQIGVDEIVKLSFNLKQKQVKNKSSYNKIDNK--HFFRKGEVRDWNLYTSEMI 333
Db 214 DPAELVDRIIQTSTFOEMKKNPCTNYSMLPETMIDLKVSFPMRKGIVGDWRNHPPEARL 273
Qy 334 KLE 337
Db 274 ERPE 277

RESULT 14
SUAR_MOUSE
ID SUAR_MOUSE STANDARD; PRT; 291 AA.
AC P52840;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Aryl sulfotransferase (EC 2.8.2.1) (Phenol sulfotransferase) (PST-1)
DE (Sulfokinase) (Phenol/aryl sulfotransferase) (ST1A4).
GN SUL1A1 OR ST1A1 OR STP1 OR STP.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
CX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=93144346; PubMed=8424956;
RA Kong A.-N.T., Ma M., Tao D., Yang L.;
RT "Molecular cloning of cDNA encoding the phenol/aryl form of
sulfotransferase (mStp1) from mouse liver.";
RL Biochim. Biophys. Acta 1171:315-318 (1993).
CC -!- FUNCTION: Catalyzes O-sulfation of phenols, N-O-sulfation of
minoxidil and tyrosine esters (By similarity).
CC -!- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + a phenol =
adenosine 3',5'-bisphosphate + an aryl sulfate.
CC -!- SUBUNIT: Homodimer (By similarity).
CC -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -!- SIMILARITY: Belongs to the sulfotransferase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch)
CC
CC EMBL; L02331; -; NOT_ANNOTATED_CDS.
CC HSP; P50224; ICMJ.
CC MGD; MGI:102896; Sult1a1.
CC InterPro; IPR000863; Sulfotransferase.
CC Pfam; PF00685; Sulfotransfer; 1.
CC ProDom; PD001218; Sulfotransferase; 1.
KW Transferase; Steroid metabolism.
FT BINDING 255 AA; 261 MW; AD07187DFASID96 CRC64;
SQ SEQUENCE 291 AA; 33974 MW; AD07187DFASID96 CRC64;

Query Match 12.1%; Score 248.5; DB 1; Length 291;
Best Local Similarity 27.6%; Pred. No. 1.2e-10;
Matches 81; Conservative 42; Mismatches 121; Indels 49; Gaps 10;

Qy 48 YSLEPNDYMGDTWRLYKGFQMGVLPVGMAFEDNFKARETDIILTLPKAGTTWKAL 107
Db 11 VKGIPLIKYFAETMEQLQ-----NFTAWPDDVLISITPKSGSTNMWSEI 53
```

MGD; MGI:98431; Ste.  
DR InterPro: IPR000863; Sulfotransferase.  
DR Pfam: PF00685; Sulfotransferase; 1.  
DR ProDom: PD001218; Sulfotransferase; 1.  
KW transferase; steroid-binding; 3D-structure.  
FT BINDING 259 265  
FT CONFLICT 113 113  
FT HELIX 8 11  
FT STRAND 12 15  
FT TURN 16 17  
FT STRAND 18 21  
FT HELIX 22 24  
FT TURN 25 27  
FT STRAND 28 32  
FT TURN 33 33  
FT STRAND 38 39  
FT TURN 41 45  
FT STRAND 47 48  
FT HELIX 51 62  
FT TURN 63 64  
FT STRAND 75 78  
FT TURN 81 81  
FT STRAND 82 83  
FT STRAND 85 86  
FT TURN 87 89  
FT STRAND 90 91  
FT HELIX 92 97  
FT TURN 98 98  
FT STRAND 104 107  
FT HELIX 111 113  
FT STRAND 116 120  
FT TURN 121 122  
FT STRAND 124 129  
FT HELIX 132 145  
FT STRAND 146 146  
FT TURN 147 148  
FT HELIX 155 164  
FT TURN 165 165  
FT STRAND 168 169  
FT HELIX 172 182  
FT TURN 183 184  
FT STRAND 186 187  
FT TURN 188 192  
FT HELIX 193 198  
FT STRAND 200 210  
FT TURN 211 212  
FT HELIX 217 226  
FT TURN 227 227  
FT HELIX 229 234  
FT TURN 236 242  
FT STRAND 245 247  
FT TURN 248 248  
FT HELIX 250 252  
FT TURN 263 266  
FT STRAND 267 267  
FT TURN 270 284  
FT TURN 285 286  
SQ SEQUENCE 295 AA; 35590 MW; 5E58AB47952BFB1C CRC64;  
Query Match 12.1%; Score 248.5; DB 1; Length 295;  
Best Local Similarity 25.8%; Pred. No. 1.2e-10;  
Matches 83; Conservative 47; Mismatches 131; Indels 61; Gaps 12;  
QY 31 EQEEDSKMKRYREIVSSLPNDYKGDTRLYKGFQWQMGYLVPGIMAFEDFKARETD 90  
DB 2 ETSMEYVEYGFGEFGLVMDKRFKYWEDV-----EMFLARPDD 40  
QY 91 ILITLTPKAGTWTWKALTFAILTR-DVNHPSSTPHLLFFNPHSCVQNLVYMGRENTM 149  
DB 41 LVIAVYPSGTTWSEVVMYIKSGDVEKCKEDA----IFN-----RIPYLECRNEDLI 90  
QY 150 PDLML--NESRPLFAGHIPYSLLPASVLYKSGTKLINISRNKSTFFYSFWEFGNLI---- 203

108 TFAL-----LTRDVNHPSPTHLFFNPHSCVQNLVYMGRENTMDDLMLNE--SPR 160  
54 MDMIYQGGKLDKCGRAFYARIPFLF-----SC-----PGVPPGLETTLKETAPR 99  
161 LFAGHIPYSLLPASVLYKSGTKIINISRNKSTFFVFWKFGNL--INPDKLLDLEKSVDF 218  
100 IKTHPLPLSLFQSLDDQKIKVIYVARNKDVVSYNFKVWAKLHDDP-GTWESFLENF 158  
219 ASGISFCQPEWFOAFTNAASTNLSLLSYEMLEKPEVNVKLAFFMCGGTFDDDEEK 278  
159 MDGKVSYSQVQHVXEMELRTHP-VLYLFYEDMKENPKREIKKILEFLGRSLPEE--- 214  
279 QGVDEIVKLCFNDLKNQVKNKSGSYNSKIDNK--HFFRKGEVRDWANYLT 329  
215 --TVDLIVHNSFKCKENPMANYTITPEVNDHTIYFFMEKGTIGDWKNTFT 265  
RESULT 15  
SUOT\_MOUSE STANDARD; PRT; 295 AA.  
AC P49951;  
DT 01-OCT-1996 (Rel. 34, Created)  
DT 15-JUL-1999 (Rel. 38, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Estrogen sulfotransferase, testis isoform (EC 2.8.2.4)  
DE (Sulfotransferase, estrogen-preferring).  
GN STE.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/KSJ-DB/DB; TISSUE=Testis;  
RX MEDLINE=95269690; PubMed=7750469;  
RA Song W.-C., Moore R., McLachlan J.A., Negishi M.;  
RT "Molecular characterization of a testis-specific estrogen  
sulfotransferase and aberrant liver expression in obese and  
diabetogenic C57BL/KSJ-db/db mice.";  
RL Endocrinology 136:2477-2484 (1995).  
RN [2]  
RP X-RAY CRYSTALLOGRAPHY (1.6 ANGSTROMS), AND REVISION TO 113.  
RC STRAIN=C57BL/KSJ-DB/DB; TISSUE=Testis;  
RX MEDLINE=98025056; PubMed=9360604;  
RA Kakuta Y., Pedersen L.G., Carter C.W., Negishi M., Pedersen L.C.;  
RT "Crystal structure of estrogen sulphotransferase.";  
RL Nat. Struct. Biol. 4:904-908 (1997).  
CC -1- FUNCTION: Sulfation of estradiol and estrone. May control the  
level of the estrogen receptor by sulfurylating free estradiol.  
CC -1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + estrone =  
adenosine 3',5'-bisphosphate + estrone 3-sulfate.  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.  
CC -1- TISSUE SPECIFICITY: Testis and at very low level in the  
placenta.  
CC -1- MISCELLANEOUS: Abnormal high expression in liver in obese and  
diabetogenic C57BL/KSJ-DB/DB strain mice. Female > male. Normal  
level in liver.  
CC -1- SIMILARITY: Belongs to the sulfotransferase family.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; S78182; AAB34320.1; -  
DR PDB; 1AQY; 28-OCT-98.  
DR PDB; 1AOY; 28-OCT-98.  
DR PDB; 1BO6; 10-AUG-99.



Db 91 NGIKQLKESEPRIVKTHLPKLLPASFWKCKMILCRNAKDVAVSYFLLMITSYP 150  
QY 204 NPKLLDLEKSYDIFASGISFCGPEWNPQAEFTNAASTNSNLLLSYEEMLEKPVNTVK 263  
Db 151 NPKSFSEF---VEKFMQGVPIYG-SWYDHVKAWWEXKNSRVLFMFYEDMKEDIRREVVK 206  
QY 264 LAEFMCGGFTDBBEKQGI VDEIVKLCSDNLRNQOVNKGSSY-----NSKIDNKH 315  
Db 207 LIEFL-----ERKPSAELVDRIIQTSTFOEMKNP-----STNYTWPEEMNQV--SPF 255  
QY 316 FRKGEVRDWANYLTSEMICKLE 337  
Db 256 MRKGIIGDWKNHFFPEALRERFD 277

Search completed: March 20, 2004, 01:38:51  
Job time : 12 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 20, 2004, 01:34:18 ; Search time 35 Seconds

(without alignments)  
3524.789 Million cell updates/sec

Title: US-09-854-122-16

Perfect score: 2059

Sequence: 1 TRGITGIAVACLPLINAGIL.....CLLSYSFVINNFILLKKKK 391

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPREMBL 25:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp Vertebrate:\*

14: sp\_unclassified:\*

15: sp\_virus:\*

16: sp\_bacteriap:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Query	Score	Match	Length	DB ID	Description
1	616	29.9	329	10	Q9M1V1	Q9M1V1 arabidopsis
2	602	29.2	323	10	Q9M1V2	Q9M1V2 arabidopsis
3	578.5	28.1	331	10	Q9FX56	Q9FX56 arabidopsis
4	572.5	27.8	331	10	Q8LEH6	Q8LEH6 arabidopsis
5	567	27.5	346	10	Q9FZ80	Q9FZ80 arabidopsis
6	562.5	27.3	338	10	Q9C9D0	Q9C9D0 arabidopsis
7	561.5	27.3	338	10	Q8LB09	Q8LB09 arabidopsis
8	558.5	27.1	350	10	Q8LA18	Q8LA18 arabidopsis
9	558.5	27.1	350	10	Q9C9C9	Q9C9C9 arabidopsis
10	555.5	27.0	359	10	Q8LSA7	Q8LSA7 arabidopsis
11	551.5	26.8	359	10	Q9FLA5	Q9FLA5 arabidopsis
12	549.5	26.7	359	10	Q940P4	Q940P4 arabidopsis
13	548.5	26.6	351	10	Q8RV79	Q8RV79 arabidopsis
14	546	26.5	325	10	Q82410	Q82410 brassica na
15	545	26.5	331	10	Q9ZPQ6	Q9ZPQ6 arabidopsis
16	544.5	26.4	326	10	Q9FZ91	Q9FZ91 arabidopsis

17	536.5	26.1	351	10	Q9FX55	Q9FX55 arabidopsis
18	533	25.9	334	10	Q8LMK8	Q8LMK8 oryza sativ
19	529.5	25.7	323	10	Q82408	Q82408 brassica na
20	528.5	25.6	332	10	Q84270	Q84270 oryza sativ
21	527	25.6	324	10	Q9ZFP5	Q9ZFP5 arabidopsis
22	525.5	25.5	333	10	Q82330	Q82330 arabidopsis
23	524	25.4	358	10	Q7XMT7	Q7XMT7 oryza sativ
24	517.5	25.1	346	10	Q7XKJ3	Q7XKJ3 oryza sativ
25	510.5	24.8	340	10	Q9FLA6	Q9FLA6 arabidopsis
26	510.5	24.8	347	10	Q8GZ53	Q8GZ53 arabidopsis
27	500.5	24.3	324	10	Q82409	Q82409 brassica na
28	499	24.2	331	10	Q9FG94	Q9FG94 arabidopsis
29	486.5	23.6	314	10	Q9STQ6	Q9STQ6 arabidopsis
30	446.5	21.7	346	10	Q8LA52	Q8LA52 oryza sativ
31	407	19.8	273	10	Q8RUC1	Q8RUC1 arabidopsis
32	403.5	19.6	287	10	Q84269	Q84269 oryza sativ
33	356	17.3	250	10	Q84NE0	Q84NE0 triticum mo
34	303	14.7	295	6	Q95MF8	Q95MF8 sus scrofa
35	278.5	13.5	301	6	Q46640	Q46640 oryctolagus
36	273.5	13.3	304	13	Q7ZYH0	Q7ZYH0 xenopus lae
37	273	13.3	299	13	Q7T1C8	Q7T1C8 brachydanio
38	270	13.1	316	11	Q8C301	Q8C301 mus musculus
39	269	13.1	301	13	Q7T1C7	Q7T1C7 brachydanio
40	268.5	13.0	301	13	Q7T2V2	Q7T2V2 brachydanio
41	266.5	12.9	307	13	Q90WR6	Q90WR6 gallus gall
42	264.5	12.8	294	13	Q7S293	Q7S293 xenopus lae
43	262	12.7	301	13	Q7ZUS4	Q7ZUS4 brachydanio
44	261	12.7	291	11	Q91X36	Q91X36 mus musculus
45	259	12.6	294	4	Q8N6X5	Q8N6X5 homo sapien

## ALIGNMENTS

RESULT 1

Q9M1V1 PRELIMINARY; PRT; 329 AA.

AC Q9M1V1

DT 01-OCT-2000 (TRENBLrel. 15, Created)

DT 01-OCT-2000 (TRENBLrel. 15, Last sequence update)

DT 01-JUN-2003 (TRENBLrel. 24, Last annotation update)

DE Sulfotransferase-like protein.

OS Arabidopsis thaliana (Mouse-ear cross).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC euroids II; Brassicales; Brassicaceae; Arabidopsis.

OX NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Jordan N., Bangert S., Wiedemann R., Voss H., Unseld M., Mewes H.W.,

RA Lemcke K., Mayer K.F.X., Quetier F., Salanoubat M.,

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA EU Arabidopsis sequencing project;

RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.

DR EMBL; AL138649; CAB72146.1; "

DR PIR; T47448; T47448.

DR HSSP; P50224; 1CUM.

DR GO; GO:0008146; F:sulfotransferase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR InterPro; IPR000863; Sulfotransferase.

DR Pfam; PF00685; Sulfotransferase; 1.

DR ProDom; PD001218; Sulfotransferase; 1.

KW Transferase.

SQ SEQUENCE 329 AA; 38059 MW; 9D04227A483A630C CRC64;

Query Match 29.9%; Score 616; DB 10; Length 329;

Best Local Similarity 41.2%; Pred. NO. 1.2e-34;

Matches 132; Conservative 62; Mismatches 100; Indels 26; Gaps 9;

QY 29 KNEQKEEDSKMYKRYRIVSSLPN-DYWGDTMLYKGFQMGVLPVGINAFEDNFKAR 87

```
Db 11 RNDELSEES-----KTLISLPSDKNGTGVNVCKYQGCWYTPPILQGLVNFQKNFQ 63
Qy 88 ETDIIITLTPKAGTTWKALTALITRDVNHPS-SPTHELLFFNPHSCVQNL-ELYMGR 145
Db 64 DTDIIIVASFPKGTWKLKALTALVRS-KHSHDDHLLSDNPAVLSPLSEMLYLCS 122
Qy 146 ENTMPDMLNTPRIFAGHYPVSLPASVLKSGTKINISNRKSTFVSWFKF----- 199
Db 123 EN--PDLTKFSSSRLLFTMPSHLTQELGKSTCKIVTMSRNVKDTLVSYWHFFCKQT 180
Qy 200 -GNLNPDKLLDLKSVDFASGISFCGPEWFOAEFTNAASTNSN-LLLLSYEMLKRP 257
Db 181 DNIIIS-----SVEDTFEFCRGWFFGFWHVLVSYNGSLEDPNHVLFMKEEMKEP 235
Qy 258 VENVKKLAFWGGTDDDEKQIVDEIVKLSFNLKQVKNKGSSNSKIDNKHFFR 317
Db 236 REQIKRLAEFLGCLFTKEEESGLVDEIIDLSLRNLSSLEINKTKLHSTGRENKTFR 295
Qy 318 KGEVDWANYLTSEMICKLE 337
Db 296 KGEVGDKWNYLTPEMENKID 315

RESULT 2
Q9M1V2 PRELIMINARY; PRT; 323 AA.
AC Q9M1V2;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-JUN-2003 (TREMBlrel. 15, Last sequence update)
DE Sulfotransferase-like protein.
GN T14D3.10.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv, Columbia;
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Alcafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chou J., Choi E., Gonzalez A.,
RA Hwang B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharly N., Pham P., Sakano H., Shinn P.,
RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC011810; AAC09547.1; -
DR HSP; P50224; 1CJM.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
SQ SEQUENCE 323 AA; 37501 MW; 00F9259E9CCF6B389 CRC64;

Query Match 29.2%; Score 602; DB 10; Length 323;
Best Local Similarity 42.6%; Pred. No. 1.1e-33;
Matches 127; Conservative 51; Mismatches 112; Indels 8; Gaps 6;

Qy 45 REIVSSLPNS-DYWGDTMRLYKGFQWQYLVPGIMAFEDNFKARETDIITLTPKAGTTW 103
Db 15 KTLISLPSDKNGTGVNVCKYQGCWYTPPILQGLVNFQKNFQ 74
Qy 104 TKALTFAITRDVNHPSPTHTLFFNPHSCVQNL-ELYMGRNTMPDMLNTPRIF 162
Db 75 LKALTFAITLHRSKQSHDDHLLSNPHVLPVFEIDLYIRSEN--PDLTKFSSSRLLF 132
Qy 163 AGHIPVSLPASVLKSGTKINISNRKSTFVSWFKFNGNLINPDK--LLDLKSVDFEAS 220
Db 133 STVPSHTLQELGKSTCKIVISNRVNDTLVSIWHFFTKQTKDEKISFDFEFNFCR 192
```

```
Qy 221 GISFCGPEWFOAEFTNAASTNSN-LLLLSYEMLKRPVENVKKLAEFMCGGTFDDDEKQ 279
Db 193 GVSIFGFWHVLVSYNGSLEDPNHVLFMKEEMKAEPRDQIKKFAEFLGCPFTKEEBS 252
Qy 280 GIYDEIVKLSFNLKQVKNKGSSNSKIDNKHFFRKEVEPDWANYLTSEMICKLE 337
Db 253 GSVDEIIDLSLRNLSSLEINKTG-KLNSGRENKMFRRKGEVGDWQNYLTPEMENKID 309

RESULT 3
Q9FX56 PRELIMINARY; PRT; 331 AA.
AC Q9FX56;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)
DE T6J4.16 protein.
GN T6J4.16.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=cv, Columbia;
RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,
RA Alcafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,
RA Buehler E., Chao Q., Chin C., Chou J., Choi E., Gonzalez A.,
RA Hwang B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,
RA Lenz C., Liu A., Liu S., Mukharly N., Pham P., Sakano H., Shinn P.,
RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.;
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AC011810; AAC09547.1; -
DR HSP; P50224; 1CJM.
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
SQ SEQUENCE 331 AA; 37709 MW; 45DC8565C780E324 CRC64;

Query Match 28.1%; Score 578.5; DB 10; Length 331;
Best Local Similarity 40.2%; Pred. No. 4.8e-32;
Matches 125; Conservative 60; Mismatches 117; Indels 9; Gaps 7;

Qy 31 EOEKEEDSKMYKYREIVVSLPSN-DYWGDTMRLYKGFQWQYLVPGIMAFEDNFKARET 89
Db 12 EEEEEENQSEETKSLISLPSIDCSGTYKLYQCYDKDILQAILNFKNFQPOET 71
Qy 90 DIILTLPKAGTTWKALTALITRDVNHPSPTHTLFFNPHSCVQNL-ELYMGRNT 148
Db 72 DIIVASFPKSGTTWKLKALTALVRSKH--TSDNHPHLLTNPHLVPYLEDLTL--KSS 127
Qy 149 MPDMLNTPRIFAGHYPVSLPASVLKSGTKINISNRKSTFVSWFKFNGNLINPDK 207
Db 128 KPDLTKLSPSSPRFLSTHMGFDALKVPKESPKIVVYVCRNVKDVLSLWCFENSMSEN 187
Qy 208 LLDLEKSVDFASGISFCGPEW-NFAEFTNAASTNSNILLISYEMLKRPVENVKLAE 266
Db 198 NLSLEALFESLCSGVNLCGLWENLVGYWRGSLDPKHVLFRLYBELKTEPRVQIKLAE 247
Qy 267 FMGCGTDDDEKQIVDEIVKLSFNLKQVKNKGSSNSKIDNKHFFRKEGEVDW 326
Db 248 FLDGPFKEEDSGVDKILELCSLRNLSSLEINKTG-SLSEGVSPKSPFRKGEVGDWKS 306
Qy 327 YLTSEMICKLE 337
Db 307 YMTPEMENKID 317

RESULT 4
Q9LEH6
```

ID Q9LEH6 PRELIMINARY; PRT; 331 AA.  
 AC Q9LEH6;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Steroid sulfotransferase, putative.  
 OS Arabidopsis thaliana (Mouse-ear cress).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
 RA Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,  
 RA Buehler E., Chao Q., Chin C., Chlou J., Choi E., Gonzalez A.,  
 RA Hong B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,  
 RA Lenz C., Liu A., Liu S., Mukharasy N., Pham P., Sakano H., Shinn P.,  
 RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.,  
 RN Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=sv. Columbia;  
 RC STRAIN=sv. Columbia;  
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,  
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,  
 RA Hayashizaki Y., Shinozaki K.,  
 RA "Arabidopsis thaliana full-length cDNA";  
 RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,  
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,  
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,  
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,  
 RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.,  
 RT "Arabidopsis Open Reading Frame (ORF) Clones";  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC026238; AAF98415.1; -;  
 DR EMBL; AK117408; BAC42075.1; -;  
 DR EMBL; BT005398; AAC63818.1; -;  
 DR PIR; E86319; E86319.  
 DR HSSP; P50224; 1CJM.  
 DR GO; GO:0008146; F:sulfotransferase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR InterPro; IPR008663; Sulfotransferase.  
 DR Pfam; PF00685; Sulfotransfer; 1.  
 DR ProDom; PD001218; Sulfotransferase; 1.  
 DR Transferrase.  
 KW SEQUENCE 331 AA; 37709 MW; 1AB5475049C64AE9 CRC64;  
 Query Match 27.8%; Score 572.5; DB 10; Length 331;  
 Best Local Similarity 39.5%; Pred. No. 1.3e-31;  
 Matches 123; Conservative 61; Mismatches 118; Indels 9; Gaps 7;  
 QY 31 EQEKEDSKMYKRYREIVSSLPN-DYGDGTWELYGFMQGLVPGIMAFEDNFKARET 89  
 DB 12 EEEEEENQSETKLISLSDIDCSGTLKYQGCYWDYDILQALINFNKNQFQPT 71  
 QY 90 DILTLPKAGTTWKALTFALTIRDVNHPSSPTPLFPNPHSCVQNYLEY-LYMGRENT 148  
 DB 72 DIIVASPPKSGTTWLKALTFALAQGRKH--TSDNHPLLTHNPHLVPYLELDLYL--KSS 127  
 QY 149 MPDLML-NESPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFTVSPKFGNLIINPK 207  
 DB 128 KPLDTKLPPSSPLRFTSMGFDALKVPLKESPKIVYVCKNVNDVLISLWCFENSMG 187  
 QY 208 LLDLEKSVDFASGIGCPGPEW-NFQAEFTNAASTNSNLLLSYEEMLEKPVENVKLA 266  
 DB 188 NLSLEALFESLCSGVNLCGLWENLVGYMRGSLDPKHVFLRYEELKTEPRVQIKRLAE 247  
 QY 267 FMCGGTDDDEKQIGVDEIVKLCSFONLKNQVKNKSSVNSKIDNKHFRKGEVDRWAN 326  
 DB 248 FLDCPPFTKEEDSGVDKILELCSRLNLSGLEINKTG-SLSEGVSPKSPFRKGEVDWKS 306  
 QY 327 YLTSEMIKKLE 337  
 DB 307 YMTPEMNKID 317  
 RESULT 5  
 Q9FZ80  
 ID Q9FZ80 PRELIMINARY; PRT; 346 AA.  
 AC Q9FZ80;  
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Putative flavonol sulfotransferase (Hypothetical protein) (Putative  
 DE Flavonol 4'-sulfotransferase).  
 GN F25116.7 OR AT1G18590/F25116.11 OR AT1G18590.  
 OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
 OC eurosids II; Brassicales; Brassicaceae; Arabidopsi.  
 OX NCBI\_TaxID=3702;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Federspiel N.A., Palm C.J., Conway A.B., Conn L., Hansen N.F.,  
 RA Altafi H., Nguyen M., Lam B., Southwick A., Miranda M., Brooks S.,  
 RA Buehler E., Chao Q., Chin C., Chlou J., Choi E., Gonzalez A.,  
 RA Hong B., Johnson-Hopson C., Khan S., Kim C., Koo T., Lee J.M.,  
 RA Lenz C., Liu A., Liu S., Mukharasy N., Pham P., Sakano H., Shinn P.,  
 RA Toriumi M., Vaysberg M., Yu G., Ecker J., Theologis A., Davis R.W.,  
 RN Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.  
 RL [2]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=sv. Columbia;  
 RC STRAIN=sv. Columbia;  
 RA Seki M., Iida K., Satou M., Sakurai T., Akiyama K., Ishida J.,  
 RA Nakajima M., Enju A., Kamiya A., Narusaka M., Carninci P., Kawai J.,  
 RA Hayashizaki Y., Shinozaki K.,  
 RA "Arabidopsis thaliana full-length cDNA";  
 RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RA Yamada K., Chan M.M., Chang C.H., Dale J.M., Hsuan V.W., Lee J.M.,  
 RA Onodera C.S., Quach H.L., Tang C., Toriumi M., Wong C., Wu H.C.,  
 RA Yu G., Yuan S., Carninci P., Chen H., Cheuk R., Hayashizaki Y.,  
 RA Ishida J., Jones T., Kamiya A., Kawai J., Kim C.J., Narusaka M.,  
 RA Nguyen M., Palm C.J., Sakurai T., Satou M., Seki M., Shinn P.,  
 RA Southwick A., Tripp M.G., Wu T., Shinozaki K., Davis R.W., Ecker J.R.,  
 RA Theologis A.,  
 RT "Arabidopsis Open Reading Frame (ORF) Clones";  
 RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC026238; AAF98415.1; -;  
 DR EMBL; AK117408; BAC42075.1; -;  
 DR EMBL; BT005398; AAC63818.1; -;  
 DR PIR; E86319; E86319.  
 DR HSSP; P50224; 1CJM.  
 DR GO; GO:0008146; F:sulfotransferase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR InterPro; IPR008663; Sulfotransferase.  
 DR Pfam; PF00685; Sulfotransfer; 1.  
 DR ProDom; PD001218; Sulfotransferase; 1.  
 DR Hypothetical protein; Transferase.  
 KW SEQUENCE 346 AA; 39912 MW; 84208C67260FD688 CRC64;  
 Query Match 27.5%; Score 567; DB 10; Length 346;  
 Best Local Similarity 39.6%; Pred. No. 3.2e-32;  
 Matches 137; Conservative 56; Mismatches 123; Indels 30; Gaps 11;  
 QY 23 EKCFGKNEQKEBDSKMYKRYREIVSSLPNDYW--GDTMRLYKGFQMGVLPVGINAF 80  
 DB 17 ELASSSPSEFKNQ-----KHVQIIATLPHKQGRPKDPVEYGGHWWLQPLLEGLHA 71  
 QY 81 EDNPKARETDILTLPKAGTTWKALTFALTIR--DVNHPSSPTPLFPNPHSCVQN 137  
 DB 72 QKFRARNDFVCYSYPTGTGTWLKALTFALANKSFDVS-----TNPLLRNPHFVPY 126  
 QY 138 LEVLYMCRENTMPDLML-NESPRLFAGHIPYSLLPASVLKSGTKIINISNRKSTFTVSF 196  
 DB 127 IEIDF---PFFPSVDVLKDSGNTLFTSTHPIYDLPSVVKSGCKIVVWRDPKDTFVM 182  
 QY 197 WKFGNLIINPKD--LLDLEKSVDFASGIGCPGPEW-NFQAEFTNAASTN-SNLLLSSEEM 253  
 DB 183 WTFPAKERSQQGPVVSIEEAFDKYCGLSAYGPDYLDHVLGYWKAYQANPDQILFLKYETM 242  
 QY 254 LEKPVENVKLAEPFGCGFTDDEEKQIGVDEIVKLCSFONLKNQVKNKSSVNSK--I 310  
 DB 243 RADPLPYVRLAEFGVGYFTKEEBEGNVKVKLCSFETLKNLEANKGEKDRPAVY 302  
 QY 311 DNKHFRKGEVDRWANYLTSEMIKKLETAGKINESEK--HLLSNKN 354  
 DB 303 ANSAYFRKGVGDQVNLTPMVARID--GLMEBKFKGTGFLSSKS 346

```

33 EKEDSKMYKRYREIVSSLPNDYW--GDTWRLYKGFQWQGYLVPGIMAFEDNFKARETD 90
14 ELTEFEKTQKQYQDFIATLPKSKGWRPDEILTYQGGHWOECLEGLFHAKDHFPEARPTD 73
91 IILTLPKAGTTWTWKALTFAITRDVNHPSPTPLFFNPHSCVQNLVLYMGRENTMP 150
74 FLVCSYPTGTTWLTALTYAIVNR--SRYDDAANPLKRNPHFVPYVEIDFA---FYP 127
151 LDMLNESPR-LFAGHIPYSLLPASVLKSGTKINISRNKSTFTSVKFGNGLINPDK-- 207
128 TVDVLQDRKNPLFSTHENGLLPDSIVNSGCKMYIWRDPKDTFISMTWTF---LHKESQ 184
208 ---LLDLEKSDVIDFASG:SPGCPENFOAERTNAASTN-SNLLLSYEEMLEKVENYVK 263
185 EQQLASLESDSPDMCKGLSVGVYLDVHLYGKAYQENPDRILFLRYETMEANLPLFVKR 244
264 LAEFMGCGFTDDEKQGVDEIVKLCSPDNLKNOQVKNKSSYNSK---INDKHFFRKE 320
245 LAEFMGYGTDEEEENGVAEKVKLCSPFTLNLEANKGKERDRPARYANSAYFRKKG 304
321 VRDMANYLTSSEMIKKELETAGKINSEKH--LISNKN 354
305 VQDMANYLTPMAARID--GLVEEKFKDTGLLQHDN 338

RESULT 7
O8LB09 PRELIMINARY; PRT; 338 AA.
ID O8LB09 AC
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Putative flavonol sulfotransferase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY087493; AAW65036.1;
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 338 AA, 39233 MW, ED9A7E0E906866F3 CRC64;

Query Match 27.3%; Score 561.5; DB 10; Length 338;
Best Local Similarity 38.4%; Pred. No. 7.4e-31;
Matches 129; Conservative 59; Mismatches 123; Indels 25; Gaps 10;

OY 33 EKEDSKMYKRYREIVSSLPNDYW--GDTWRLYKGFQWQGYLVPGIMAFEDNFKARETD 90
DB 14 ELTEFEKTQKQYQDFIATLPKSKGWRPDEILTYQGGHWOECLEGLFHAKDHFPEARPTD 73
OY 91 IILTLPKAGTTWTWKALTFAITRDVNHPSPTPLFFNPHSCVQNLVLYMGRENTMP 150
DB 74 FLVCSYPTGTTWLTALTYAIVNR--SRYDDAANPLKRNPHFVPYVEIDFA---FYP 127
OY 151 LDMLNESPR-LFAGHIPYSLLPASVLKSGTKINISRNKSTFTSVKFGNGLINPDK-- 207
DB 128 TVDVLQDRKNPLFSTHENGLLPDSIVNSGCKMYIWRDPKDTFISMTWTF---LHKESQ 184
OY 208 ---LLDLEKSDVIDFASG:SPGCPENFOAERTNAASTN-SNLLLSYEEMLEKVENYVK 263
DB 185 EQQLASLESDSPDMCKGLSVGVYLDVHLYGKAYQENPDRILFLRYETMEANLPLFVKR 244
OY 264 LAEFMGCGFTDDEKQGVDEIVKLCSPDNLKNOQVKNKSSYNSK---INDKHFFRKE 320
DB 245 LAEFMGYGTDEEEENGVAEKVKLCSPFTLNLEANKGKERDRPARYANSAYFRKKG 304
OY 321 VRDMANYLTSSEMIKKELETAGKINSEKH--LISNKN 354
DB 305 VQDMANYLTPMAARID--GLVEEKFKDTGLLQHDN 338

RESULT 7
O8LB09 PRELIMINARY; PRT; 338 AA.
ID O8LB09 AC
DT 01-OCT-2002 (TREMBLrel. 22, Created)
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TREMBLrel. 24, Last annotation update)
DE Putative flavonol sulfotransferase.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volkovskiy N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation.";
RL Genome Biol. 0:0-0 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY087493; AAW65036.1;
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 338 AA, 39233 MW, ED9A7E0E906866F3 CRC64;

Query Match 27.3%; Score 561.5; DB 10; Length 338;
Best Local Similarity 38.4%; Pred. No. 7.4e-31;
Matches 129; Conservative 59; Mismatches 123; Indels 25; Gaps 10;

OY 33 EKEDSKMYKRYREIVSSLPNDYW--GDTWRLYKGFQWQGYLVPGIMAFEDNFKARETD 90
DB 14 ELTEFEKTQKQYQDFIATLPKSKGWRPDEILTYQGGHWOECLEGLFHAKDHFPEARPTD 73
OY 91 IILTLPKAGTTWTWKALTFAITRDVNHPSPTPLFFNPHSCVQNLVLYMGRENTMP 150
DB 74 FLVCSYPTGTTWLTALTYAIVNR--SRYDDAANPLKRNPHFVPYVEIDFA---FYP 127
OY 151 LDMLNESPR-LFAGHIPYSLLPASVLKSGTKINISRNKSTFTSVKFGNGLINPDK-- 207
DB 128 TVDVLQDRKNPLFSTHENGLLPDSIVNSGCKMYIWRDPKDTFISMTWTF---LHKESQ 184
OY 208 ---LLDLEKSDVIDFASG:SPGCPENFOAERTNAASTN-SNLLLSYEEMLEKVENYVK 263
DB 185 EQQLASLESDSPDMCKGLSVGVYLDVHLYGKAYQENPDRILFLRYETMEANLPLFVKR 244
OY 264 LAEFMGCGFTDDEKQGVDEIVKLCSPDNLKNOQVKNKSSYNSK---INDKHFFRKE 320
DB 245 LAEFMGYGTDEEEENGVAEKVKLCSPFTLNLEANKGKERDRPARYANSAYFRKKG 304
OY 321 VRDMANYLTSSEMIKKELETAGKINSEKH--LISNKN 354
DB 305 VQDMANYLTPMAARID--GLVEEKFKDTGLLQHDN 338

```



```

SQ SEQUENCE 350 AA; 40465 MW; AB68B7477262EB57 CRC64;
Query Match 27.1%; Score 558.5; DB 10; Length 350;
Best Local Similarity 39.0%; Pred. No. 1.2e-30;
Matches 129; Conservative 59; Mismatches 120; Indels 23; Gaps 9;
QY 29 KNEQKEEDSKMYRYRIVSLPSNDYWGDMRL--YKGFQWQYLVPGIMAFEDNKA 86
D 22 ETKTSTTEFKNQKQYQDLSTTFPEKGRWPKREPLIEYGGYVWLPSSLEGCIHAQEFFQA 81
QY 87 RETDILITLTPKAGTWTWKALTFALLTRDYNHPSPTPLFFNPHSCVQWQVLELYMGRE 146
D 82 RPSDFLVCSYPTGTTWTKALTFALTANESRPDSS--NPLKRNPHFVPIEDF---- 135
QY 147 NTMPDMLNESPRLPAGHYPISLLPASVLSKGTKIINISNRKSTFVSPWKGNGLINP 205
D 136 PFFPEVDLKDGNLFTTHIPYELLPSVVKSGCKMYIWRPKDTFISMWTF---LHK 192
QY 206 DR-----LIDLEKSVDFISGIFCGPEWFOAEFTNAASTN--SNLLLSYEEMLEKPE 259
D 193 ERTGLGVSINLESEDFMCRGLSGVGYVYNHILAWKAYQENPDRIELFLKYETWRADPLP 252
QY 260 NVKLAEPFGCGFTDDEKQGVIVDEIVKLCFDPNLSQVQVNGSSYNSK---IDNKHFF 316
D 253 YVKSLEAPMGHGFATSEKGVYEVKVNLCSEFETLKNLEANKGEKDRDPGVYANGAYF 312
QY 317 RKGEVRDWNLYLTSEMICKLETPAGNINESEK 347
D 313 RKKGVDGWSNYLTPEMARID--GLMEKFK 341

RESULT 10
Q8L5A7 PRELIMINARY; PRT; 359 AA.
AC Q8L5A7;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Steroid sulfotransferase-like protein (At5g07010).
GN AT5G07010.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA Nguyen M., Karlin-Neumann G., Southwick A., Lam B., Miranda M.,
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,
RA Ecker J., Theologis A., Davis R.W.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,
RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;
RT "Full-length messenger RNA sequences greatly improve genome
RT annotation."
RL Genome Biol. 0:0-0(2002).
RN [3]
RP SEQUENCE FROM N.A.
RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,
RA Feldmann K.;
RT "Full-length cDNA from Arabidopsis thaliana."
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RA Kim C.J., Chen H., Cheuk R., Shinn P., Bowser L., Carninci P.,
RA Dale J.M., Hayashizaki Y., Ishida J., Jones T., Kamiya A.,
RA Karlin-Neumann G., Kawai J., Lam B., Lin J., Miranda M., Narusaka M.,
RA Nguyen M., Onodera C.S., Palm C.J., Quach H.D., Sakurai T., Satou M.,
RA
```

```

RA Seki M., Southwick A., Toriumi M., Wong C., Wu H.C., Yamada K., Yu G.,
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;
RT "Arabidopsis ORF clones."
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AY099809; AAM20660.1; -
DR EMBL; AY084999; AAM61557.1; -
DR EMBL; BT008847; AAP68286.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
KW Transferase.
SQ SEQUENCE 359 AA; 41364 MW; 04244AD987C3F104 CRC64;
Query Match 27.0%; Score 555.5; DB 10; Length 359;
Best Local Similarity 35.9%; Pred. No. 2.1e-30;
Matches 123; Conservative 69; Mismatches 140; Indels 11; Gaps 8;
QY 1 TRGITGIAVACLPLINMAGILALEKCFGKNEKEEDSKMYRYRIVSLPSNDYWGDT 60
D 3 TSSMKIPNAPISFNSCHKLELLKSGKTRDVPKAEDEGLSCFQEMLDLSLPERGWR 62
QY 61 -MRLYKGFQWQYLVPGIMAFEDNPKARETDIITLTPKAGTWTWKALTFALLTRDYNH 119
D 63 YLYFGQFWQCAKEIQAIMSFQKFSQLENDVVLATIPKSGTTLKALTFTTLNRHRDP 122
QY 120 --SSPTHPLLFNPHSCVQWQVLEY-LYMGRENTPDLDMLNESPRLPAGHYPISL 176
D 123 VASSTNHPLFTSNPHDLVFFYKLYANGD--VPDLISGL-ASPRTFATHLPGSLKET 179
QY 177 KSGTKIINISNRKSTFVSWFKGNLINPDKL--LDLEKSVDFISGIFCGPEWFOAE 234
D 180 KPGVKVYLCRNPFDTFISSWHYTNKSESVPVLLDQAFDLYCRGVIGFGPFWEHMLG 239
QY 235 F-TNAASTNSNLLLSYEEMLEKPEVNVKLAEPFGCGFTDDEKQGVIVDEIVKLCSP 293
D 240 YWRESLKRPKEKVFYFLYEDKDDIETNLKRLATLELPTFEEERAGVVKALAE 299
QY 294 LKNQVQNGSSYNSKINDKNHFFRGEVRDWNLYLTSEMICKL 336
D 300 LKLEVNKSNKSIKN-FENRFLFRKGEVSDWVNYLSPSQVERL 341

RESULT 11
Q9FL45 PRELIMINARY; PRT; 354 AA.
AC Q9FL45;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-JUN-2003 (Tremblrel. 24, Last annotation update)
DE Steroid sulfotransferase-like protein.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Columbia;
RA MEDLINE=99344145; PubMed=9679202;
RA Kaneko T., Kotani H., Nakamura Y., Sato S., Asamizu E., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. V. Sequence
RT features of the regions of 1,381,565 bp covered by twenty one
RT physically assigned P1 and TAC clones."
RL DNA Res. 5:131-145(1998).
DR EMBL; AB010697; BAB11159.1; -
DR GO; GO:0008146; F:sulfotransferase activity; IEA.
DR GO; GO:0016740; F:transferase activity; IEA.
DR InterPro; IPR000863; Sulfotransferase.
DR Pfam; PF00685; Sulfotransferase; 1.
DR ProDom; PD001218; Sulfotransferase; 1.
```

KW Transferase.  
SQ SEQUENCE 354 AA; 40887 MW; E540A1449966A36C CRC64;  
Query Match 26.8%; Score 551.5; DB 10; Length 354;  
Best Local Similarity 36.2%; Pred. No. 3.8e-30;  
Matches 122; Conservative 68; Mismatches 136; Indels 11; Gaps 8;  
QY 7 IAVACLPLIMAGILALEKFCGSKNEQKEEDSKMYKRYREIVSSLPNDYWGDT-MELYK 65  
DB 4 IPMAIPSPSMCHKLELLKGGTRDVPKAEDEGLSCFQEMDLSPKRGWRIRYLYLQ 63  
QY 66 GFWQGYLVPGIMAFENFKAETDITLTPKAGTTWTKALTPTAILTRDVNHP--SSPT 123  
DB 64 GFWQCAKEIOAIMSFQKHSLENDVWLATIPKSGTTLKALTPTILNRHFRDPVASSTN 123  
QY 124 HPLLFNPHSHSCVQNLLEY-LYMGRENTMPDLMLNESPRLFAGHPYSLPASVLSKGTKI 182  
DB 124 HPLTSPNPHDLVPFFYKLYANGD--VPDLSGL-ASPRTFATHLPFGSLKXETIEKPKVKV 180  
QY 183 INISNRKSTFTVSWFKFNGNLINPKL--LDLEKSVDFASGISFCGPEWNFQAEF-TNAA 239  
DB 181 VYLRCNPDPDTTSSWHYTNNTKSESVSPVLLDQAFDLYCRGVIGFPGFWEHMLGYWRESL 240  
QY 240 STNSNLLLSYEEMLEKPEVENVKLAERMGCGFTDDSEKQGVDEIVKLSFENLKNQOV 299  
DB 241 KRPEKVFRLRYEDLDKDIETNLKLTATFLFPFTEEEERKGVYKAIABLSFENLKKLEV 300  
QY 300 NKGSSYNKSIDNKHFRKGEVRDWAYNYLTSEMICKL 336  
DB 301 NKSNSIKN-PENRFLFRKGEVSDWNYLSPSQVERL 336  
RESULT 12  
Q940B4 PRELIMINARY; PRT; 359 AA.  
AC Q940P4;  
DT 01-DEC-2001 (TREMELrel. 19, Created)  
DT 01-DEC-2001 (TREMELrel. 19, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE AT5G07010/MOJ9.19.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Shinn P., Chen H., Cheuk R., Kim C.J., Koesema E., Meyers M.C.,  
RA Shinn J., Bowser L., Carninci P., Dale J.M., Goldsmith A.D.,  
RA Hayashizaki Y., Ishida J., Jiang P.X., Jones T., Kamiya A.,  
RA Karlin-Neumann G., Kawai J., Lam B., Lee J.M., Lin J., Liu S.X.,  
RA Miranda M., Narusaka M., Nguyen M., Onodera C.S., Palm C.J.,  
RA Pham P.K., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
RA Tang C.C., Toriumi M., Yamada K., Yamamura Y., Yu G., Yu S.,  
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
RT "Arabidopsis cdna clones."  
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
DR GO: 0008146; F:sulfotransferase activity; IEA.  
DR EMBL; AY054219; AAL06879.1;  
DR InterPro; IPR000863; Sulfotransferase.  
DR Pfam; PF00685; Sulfotransferase.  
DR ProDom; PD001218; Sulfotransferase; 1.  
SQ SEQUENCE 359 AA; 41378 MW; 061A4F79873CF44 CRC64;  
Query Match 26.7%; Score 549.5; DB 10; Length 359;  
Best Local Similarity 35.6%; Pred. No. 5.4e-30;  
Matches 122; Conservative 69; Mismatches 141; Indels 11; Gaps 8;  
QY 1 TRGTGIAVACLPLIMAGILALEKFCGSKNEQKEEDSKMYKRYREIVSSLPNDYWGDT 60  
DB 3 TSSMKSIPIWAIPTFSFMCHKLELLKGGTRDVPKAEDEGLSCFQEMDLSPKRGWRIR 62  
QY 61 -MRLYKGFQWQGYLVPGIMAFENFKAETDITLTPKAGTTWTKALTPTAILTRDVNHP 119

DB 63 VLYLFGQFWQCAKEIOAIMSFQKHSLENDVWLATIPKSGTTLKALTPTILNRHFRDP 122  
QY 120 --SSPTHLLPFNPHSCVQNLLEY-LYMGRENTMPDLMLNESPRLFAGHPYSLPASV 176  
DB 123 VASSTNPHLFTSNPHDLVPFFYKLYANGD--VPDLSGL-ASPRTFATHLPFGSLKXETIE 179  
QY 177 KSGTKIINISNRKSTFTVSWFKFNGNLINPKL--LDLEKSVDFASGISFCGPEWNFQAE 234  
DB 180 KPGVKVYVLYCRNPDPDTTSSWHYTNNTKSESVSPVLLDQAFDLYCRGVIGFPGFWEHMLG 239  
QY 235 F-TNAASTNSNLLLSYEEMLEKPEVENVKLAERMGCGFTDDSEKQGVDEIVKLSFEN 293  
DB 240 YWRESLKPEKPEKVFRLRYEDLDKDIETNLKLTATFLFPFTEEEERKGVYKAIABLSFEN 299  
QY 294 LKNQOVNKGSSYNKSIDNKHFRKGEVRDWAYNYLTSEMICKL 336  
DB 300 LKLELVNKSNIK-PENRFLFRKGEVSDWNYLSPSQVERL 341  
RESULT 13  
Q8RV79 PRELIMINARY; PRT; 351 AA.  
ID Q8RV79;  
AC Q8RV79;  
DT 01-JUN-2002 (TREMELrel. 21, Created)  
DT 01-JUN-2002 (TREMELrel. 21, Last sequence update)  
DT 01-JUN-2003 (TREMELrel. 24, Last annotation update)  
DE Putative steroid sulfotransferase.  
GN AT2G03750.  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=cv. Columbia;  
RA Lin X., Kaul S., Shea T.P., Fujii C.Y., Shen M., VanAken S.E.,  
RA Barnstead M.E., Mason T.M., Bowman C.L., Ronning C.M., Benito M.-I.,  
RA Carrera A.J., Creasy T.H., Buel C.R., Town C.D., Nieman W.C.,  
RA Fraser C.M., Venter J.C.;  
RL Submitted (FEB-2002) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Nguyen M., Karlin-Neumann G., Southwick A., Tripp M., Miranda M.,  
RA Palm C.J., Bowser L., Jones T., Banh J., Carninci P., Chen H.,  
RA Cheuk R., Chung M.K., Hayashizaki Y., Ishida J., Kamiya A., Kawai J.,  
RA Kim C., Lin J., Liu S.X., Narusaka M., Pham P.K., Sakano H.,  
RA Sakurai T., Satou M., Seki M., Shinn P., Yamada K., Shinozaki K.,  
RA Ecker J., Theologis A., Davis R.W.;  
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC006836; RAD20077.2;  
DR EMBL; AY092961; AAM12960.1;  
DR EMBL; BT002108; AAN72119.1;  
DR GO: 0008146; F:sulfotransferase activity; IEA.  
DR GO: 0008146; F:sulfotransferase activity; IEA.  
DR InterPro; IPR000863; Sulfotransferase.  
DR Pfam; PF00685; Sulfotransferase.  
DR ProDom; PD001218; Sulfotransferase; 1.  
KW Transferase.  
SQ SEQUENCE 351 AA; 40938 MW; 2B8B7F61060333FA CRC64;  
Query Match 26.6%; Score 548.5; DB 10; Length 351;  
Best Local Similarity 37.4%; Pred. No. 6.1e-30;



Matches 125; Conservative 61; Mismatches 125; Indels 23; Gaps 10;

QY 22 LEKFGS-----KNEQKE-----EDSKYKRYREIVSSLPNS-DYMGDTNRLYKGF 67

DB 8 LSLCFKSNEQIITWEASKEAHLNPMKDDNVSQETKNLITSLPSDKDFMGYGLNYKGC 67

QY 68 WQMGVLPVGIWAFEDNFKARETDILITLTPKAGTWTWKALTFALLTRDYNHPSPTPL 127

DB 68 WYFNTLQAVLDVQKHPRDITLALSLPKGTTWLKSLIPAVVHREKRGRTQTPTPL 127

QY 128 FNPNSVCQNLLEY-LYMGRENTMPDMLNESPRLFAGHIPYSLLPASVLSKSGTKIINIS 186

DB 128 LQNPDLVPFLEVELYANSQ--IPLAKYS-SPMIFSTHMLQALREATTKA-CKTVVVC 183

QY 187 NNRKSTFVSFWKFGNLINPKL--LDLEKSVDFIPASGISFCGPWFNFOAEF-TWAASNS 243

DB 184 RGIKDTFVSGMHRNMLHRTQDQATFELMFDAYCGVLLYGPYWEHVLVSYNKGSLAKE 243

QY 244 NLLLSYEEMLEKPVENKLAEFMCGFTDDEKQGVDEIVKLSFDNLKQVKNKG 303

DB 244 NVLFKRYEIIIEPRVQVKRLAEFLCEFTKEEESGVSBEILKLSLRLNLSLVNKG 303

QY 304 SSYNSKIDNKHFFRKGEVDWANYLTSEMICKLE 337

DB 304 TT-RIGVDSQVFRKGEVGDWKNHLPQWAKTFD 336

RESULT 14

OB2410 PRELIMINARY; PRT; 325 AA.

AC OB2410;

DT 01-NOV-1998 (TrEMBLrel. 08, Created)

DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Steroid sulfotransferase 3.

OS Brassica napus (Rapel).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Brassica.

OC NCBI\_TaxID=3708;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=99340019; PubMed=10409637;

RA Rouleau M., Marsolais F., Richard M., Nicolle L., Voigt B., Adam G.,

RA Varin L.;

RT "Inactivation of brassinosteroid biological activity by a salicylate-

RT inducible steroid sulfotransferase from Brassica napus.";

RL J. Biol. Chem. 274:20925-20930(1999).

DR ENBL; AF000307; AAC63113.1; -

DR PIR; T07833; T07833.

DR HSP; P50224; ICJM.

DR GO; GO:0008146; F:sulfotransferase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR InterPro; IPR000863; Sulfotransferase.

DR Pfam; PF00685; Sulfotransfer; 1.

DR ProDom; PD001218; Sulfotransferase; 1.

KW Transferase.

SQ SEQUENCE 325 AA; 36940 MW; 2DD1BCBCEE203AE1 CRC64;

Query Match 26.5%; Score 546; DB 10; Length 325;

Best Local Similarity 40.2%; Pred. No. 8.2e-30;

Matches 123; Conservative 52; Mismatches 117; Indels 14; Gaps 7;

QY 37 DSKMYKRYREIVSSLPNSDYW-GDTWELYKGFQMGVLPVPGIMAFEDNFKARETDILIT 95

DB 13 DEKLTQTRLLISLSEKGLVSIQYQFQGRWHTEALQGLITCQHFKAQSDILIT 72

QY 96 LPKAGTTWTKALTFALLTRDYNHPSPTPLFPNPHSCVQNLLEYLYMGRENTMPDMLN- 154

DB 73 NPXSGTTWLKSLVFLALINRHKFPVSSGDHPLLVTNPHELLVPFMGVVY----ESPDFDS 128

QY 155 LNESPRLFAGHIPYSLLPASVLSKGTKIINISNRKSTFVSFWKFGNLINPKL--LE 212

DB 129 LLPPFLMNTHTSHLSLPESVYKSSSCQIVYCCRNPKDMFVSLWHPGKGLAPQETADYPLE 188

QY 213 KSVDPFASGISFCGPWFNFOAEFTWAASNSN-LLLLSYEEMLEKPVENKLAEFMCG 271

DB 189 KAVEAFCCQKTAGFWDHVLLEYWYASLENPKLVFTVYEUKQTEVEVKRIAEPICG 248

QY 272 FTDDKEKQIVDEIVKLSFDNLKQVKNKGSSYNSKIDNKHFFRKGEVDWANYLTSE 331

DB 249 FTAEE-----VSEIVKLSFESLSRLVNRQKLPNG-IETNAPFRKGEIGWRDTLSES 303

QY 332 MIKKE 337

DB 304 LADAID 309

RESULT 15

Q92PQ6 PRELIMINARY; PRT; 331 AA.

AC Q92PQ6;

DT 01-MAY-1999 (TrEMBLrel. 10, Created)

DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)

DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)

DE Putative steroid sulfotransferase.

GN A2G03750

OS Arabidopsis thaliana (Mouse-ear cress).

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;

OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

OC eurosids II; Brassicales; Brassicaceae; Arabidopsis.

OC NCBI\_TaxID=3702;

RN [1]

RP SEQUENCE FROM N.A.

RA Southwick A., Karlin-Neumann G., Nguyen M., Lam B., Miranda M.,

RA Palm C.J., Theologis A., Ecker J., Davis R.W.;

RL Submitted (NOV-2000) to the EMBL/GenBank/DBJ databases.

RN [2]

RP SEQUENCE FROM N.A.

RA Haas B.J., Volfovsky N., Town C.D., Troukhan M., Alexandrov N.,

RA Feldmann K.A., Flavell R.B., White O., Salzberg S.L.;

RT "Full-length messenger RNA sequences greatly improve genome

RT annotation.";

RL Genome Biol. 0:0-0(2002).

RN [3]

RP SEQUENCE FROM N.A.

RA Brover V., Troukhan M., Alexandrov N., Lu Y.-P., Flavell R.,

RA Feldmann K.;

RT "Full-length cDNA from Arabidopsis thaliana.";

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR ENBL; AF325065; AAK17133.1; -

DR EMBL; AY087921; AAM65471.1; -

DR PIR; H84451; H84451.

DR HSP; P49891; 1AQU.

DR GO; GO:0008146; F:sulfotransferase activity; IEA.

DR GO; GO:0016740; F:transferase activity; IEA.

DR InterPro; IPR000863; Sulfotransferase.

DR Pfam; PF00685; Sulfotransfer; 1.

DR ProDom; PD001218; Sulfotransferase; 1.

KW Transferase.

SQ SEQUENCE 331 AA; 38400 MW; 22E0E0A2D8CF1154 CRC64;

Query Match 26.5%; Score 545; DB 10; Length 331;

Best Local Similarity 38.4%; Pred. No. 9.9e-30;

Matches 118; Conservative 61; Mismatches 118; Indels 10; Gaps 8;

QY 36 EDSKRYKRYREIVSSLPNS-DYMGDTWRLYKGFQMGVLPVPGIMAFEDNFKARETDILIT 94

DB 15 KDNVSQETKNLITSLPSDKDFMGYGLNYKGCWYYPNTLQAVLDVQKHPRDITLILA 74

QY 95 TLKAGTTWTKALTFALLTRDYNHPSPTPLFPNPHSCVQNLLEY-LYMGRENTMPDMLN- 153

DB 75 SLPGGTTWLKSLIPAVVHREKRGRTQTPTPLLPNPHDLVPFLEVELYANSQ--IFDLA 132

QY 154 MLNESPRLFAGHIPYSLLPASVLSKGTKIINISNRKSTFVSFWKFGNLINPKL--LDL 211

Db	133	KYS-SPMIFSTHMLQALREATTKA-CKTVVCRGIGKDFVSGWHYRNMHLRTKMDQATF	190
Qy	212	EKSVDFIPASGISFCGPENNFQAEF-TNAASTNSNLLLSYEEMLEKPEVNVKLAEPFMC	270
Db	191	ELMFDAYCRGVLLYGPYWEHVLGYWKGSLKAKENVLFMKYEEIEEPVQVKRLAEFFEC	250
Qy	271	GFTDDREKQGI VDEIVKLCSPDNLCQOVNKGSSVNSKIDNKHFFRKGEVRDWNALTS	330
Db	251	PFTKEEESGSVEEILKCLSLRNLNLEVNKNGTT-RIGVDSQVFFRKGEVGDWKNHLP	309
Qy	331	EMIKKLE	337
Db	310	QMAKTFD	316

Search completed: March 20, 2004, 01:38:31  
Job time : 38 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - nucleic search, using frame\_plus\_p2n model

Run on: March 20, 2004, 03:44:29 ; Search time 311 Seconds  
(without alignments)  
3526.461 Million cell updates/sec

Title: US-09-854-122-16  
Perfect score: 2059  
Sequence: 1 TRGITGIAVACLEPLIMAGIL.....CLLSYSFVNNFILLKKKK 391

Scoring table: BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Zgapop 6.0 , Zgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 27513289 seqs, 14931090276 residues

Total number of hits satisfying chosen parameters: 55026578

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+p2n.model -DEV=xih  
-Q=/cgn2\_1/USPTO.spool/US09854122/runat\_18032004\_100000\_3261/app\_query.fasta\_1.583  
-DB=EST -QWTF=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-USER=US09854122@cgn\_1\_2810 @runat\_18032004\_100000\_3261 -NCPU=6 -ICPU=3  
-NO MMAP -LARGEQUERY -NEG SCORES=0 -WAIT -DSFBLOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

- 1: em\_estba:\*\*
- 2: em\_esthum:\*\*
- 3: em\_estin:\*\*
- 4: em\_estmu:\*\*
- 5: em\_estov:\*\*
- 6: em\_estpl:\*\*
- 7: em\_estro:\*\*
- 8: em\_htc:\*\*
- 9: gb\_estl:\*\*
- 10: gb\_estc:\*\*
- 11: gb\_htc:\*\*
- 12: gb\_est3:\*\*
- 13: gb\_est4:\*\*
- 14: gb\_est5:\*\*
- 15: em\_estfun:\*\*
- 16: em\_estom:\*\*
- 17: em\_gss\_hum:\*\*
- 18: em\_gss\_inv:\*\*
- 19: em\_gss\_pln:\*\*
- 20: em\_gss\_vrt:\*\*
- 21: em\_gss\_fun:\*\*
- 22: em\_gss\_man:\*\*
- 23: em\_gss\_mus:\*\*
- 24: em\_gss\_pro:\*\*
- 25: em\_gss\_rod:\*\*
- 26: em\_gss\_phg:\*\*
- 27: em\_gss\_vrl:\*\*
- 28: gb\_gssl:\*\*

29: gb\_gss2:\*\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	522.5	25.4	843	29	CC701537	CC701537 OG0BJ24TV
C	510	24.8	831	28	BZ797778	BZ797778 PUFZ85TB
3	501	24.3	1137	14	CK209284	CK209284 FGAS02104
4	499.5	24.3	975	29	CG290240	CG290240 OGZAR72TV
C	485.5	23.6	990	29	CC698091	CC698091 OG0BP20TV
C	482	23.4	859	14	CF833811	CF833811 UCRCS02_0
C	471	22.9	870	29	CC691394	CC691394 OGUEY95TV
C	471	22.9	1010	29	CC629410	CC629410 OGUBW78TV
C	469.5	22.8	934	29	CG919663	CG919663 MBET65TR
10	469	22.8	950	10	BF065941	BF065941 HV_CED001
11	466.5	22.7	1118	14	CB330744	CB330744 SPST709
C	451	21.9	877	29	CC614744	CC614744 OGUGO80TV
13	436	21.2	807	29	CG196721	CG196721 PUTCM22TB
14	432	21.0	855	29	CG958535	CG958535 MBGCO3TF
15	426.5	20.7	966	29	CG289067	CG289067 OGKCR72TV
C	412	20.6	872	29	CG848241	CG848241 ZMMBB5031
17	412	20.0	819	28	BZ490661	BZ490661 BONIS44TR
18	411	20.0	769	14	CF554514	CF554514 GGN98 Inf
19	411	20.0	843	29	CG300068	CG300068 OGXF123TV
20	410	19.9	694	13	BUS34467	BUS34467 T061E09 P
21	407.5	19.8	738	14	CF418052	CF418052 UODA-PP_1
C	407	19.8	757	28	BZ062934	BZ062934 LLC02BLL
22	405.5	19.7	698	13	BUS31310	BUS31310 T020A09 P
C	403.5	19.6	717	28	BZ080154	BZ080154 11A58CL1
24	403.5	19.6	725	28	BH549521	BH549521 BORCD19TF
25	403.5	19.6	731	14	CA990082	CA990082 EST643590
26	403.5	19.6	731	14	CA990082	CA990082 OGVEY34TC
27	402	19.5	982	29	CC602680	CC602680 sat35810
28	400.5	19.5	795	14	CA799562	CA799562 sat35810
29	398.5	19.4	580	10	BE942160	BE942160 EST421739
30	396	19.2	811	28	BH733391	BH733391 BOMIU60TR
31	395	19.2	869	29	CG848310	CG848310 ZMMBB5031
C	393	19.1	1074	14	CK213074	CK213074 FGAS02497
33	392	19.0	630	14	CA990182	CA990182 EST643690
34	391.5	19.0	828	28	BH705401	BH705401 BOMBB06TF
C	391	18.9	895	29	CG352888	CG352888 OGOF115TH
36	388.5	18.9	734	28	CC347438	CC347438 OGMEB41TV
37	387.5	18.8	781	14	CA782758	CA782758 sat54511
38	386.5	18.8	624	14	CD427868	CD427868 ETH1_34 G
39	385	18.7	689	12	BG600378	BG600378 EST505273
40	383.5	18.6	656	14	CF810624	CF810624 CA327F CD
41	383	18.6	1164	14	CK171170	CK171170 FGAS04631
42	382	18.6	1098	14	CK207849	CK207849 FGAS01951
43	381.5	18.5	610	14	CF810628	CF810628 CA330F CD
44	380	18.5	786	10	BF275081	BF275081 GA_EB002
45	379.5	18.4	699	14	CD862457	CD862457 AZ01_103J

ALIGNMENTS

RESULT 1  
CC701537  
LOCUS  
DEFINITION  
OG0BJ24TV ZM.0.7.1.5 KB Zea mays genomic clone ZMMBMA034D24,  
genomic survey sequence.  
ACCESSION  
CC701537  
VERSION  
CC701537.1  
KEYWORDS  
GSS.  
SOURCE  
Zea mays  
ORGANISM  
Zea mays

CC701537 843 bp DNA linear GSS 19-JUN-2003  
OG0BJ24TV ZM.0.7.1.5 KB Zea mays genomic clone ZMMBMA034D24,  
genomic survey sequence.  
ACCESSION  
CC701537  
VERSION  
CC701537.1  
KEYWORDS  
GSS.  
SOURCE  
Zea mays  
ORGANISM  
Zea mays  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
clade; Panicoideae; Andropogoneae; Zea.

```

REFERENCE 1 (bases 1 to 843)
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,
         Resnick,A., Fraser,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,
         Citek,R.W., Nunberg,A., Robbins,D. and Lakey,N.
TITLE Consortium for Maize Genomics
JOURNAL Unpublished (2002)
COMMENT Other_GSSs: OGOBJ24TH
         Contact: Cathy Whitelaw
         TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA
         Tel: 301-838-5843
         Fax: 301-838-0208
         Email: whitelaw@tigr.org
         Seq primer: TR
         Class: sheared ends.
FEATURES
    source
        location/Qualifiers
            1..843
            /organism="Zea mays"
            /mol_type="genomic DNA"
            /strain="B73"
            /db_xref="taxon:4577"
            /clone_lib="ZM0.7 1.5 KB"
            /clone_lib="ZM0.7 1.5 KB"
            /note="Vector: pBGSK; Site 1: HincII; 0.7-1.5 kb
            methylation filtered genomic DNA library"
ORIGIN
Alignment Scores:
Pred. No.: 4,018-47 Length: 843
Score: 522.50 Matches: 111
Percent Similarity: 57.53% Conservative: 57
Best Local Similarity: 38.01% Mismatches: 99
Query Match: 25.38% Indels: 25
DB: 29 Gaps: 7
US-09-854-122-16 (1-391) x CC701537 (1-843)
QY 56 TyrTrpGlyAspThrMetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValPro 75
   ::::::::::::::::::::
Db 2 TTTTGGCTCGCTGAGTAGCATTGAAGGAGGC-----GTCCCC 40
QY 76 GlyIleMetAlaPheGluAspAsnPhaLysAlaArgGluThrAspIleLeuThrThr 95
   ::::::::::::::::::::
Db 41 GCGGTT-----CATTCTCGCTTCGAGCCGAGACCCACCGAGTGTCTCCCAAGC 91
QY 96 LeuProLysAlaGlyThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAsp 115
   ::::::::::::::::::::
Db 92 TTTCCCAAGTCGTGCACACCTGGCTCAAGGCCCTCGCTTCGCGACGCTGAAGCGTTCC 151
QY 116 ValAsnHisProSerSerProThrHisProLeuLeuPheAsnProHisSerCysVal 135
   ::::::::::::::::::::
Db 152 ACGCATCCGCGCTGTGCAGCAGCACCATCCGCTGCGCCATTGCCAACCCTCATGACTGTCTC 211
QY 136 GlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeu 155
   ::::::::::::::::::::
Db 212 AGGTTCTCGAATTGATTCAACAGCAGAGAGGATGACTCGAGGCGTCTC----- 262
QY 156 AsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerVal 175
   ::::::::::::::::::::
Db 263 ---CCGTCCCGCGTGTGTGCGCAGCACCTCCCTACTCCCTGTGTGCGCGCAGCATC 319
QY 176 Leu-----LysSerGlyThrLysIleIleAsnIleSerArgAsnArgLysSer 191
   ::::::::::::::::::::
Db 320 ACCGGCAGGAGAGCGCTCGGGGTCCGGATCGTCTACGTTTGGCGGAACCCAGGAC 379
QY 192 ThrPheValSerPheTrpLysPheGlyAsnLeuIleAsnProAspLysLeuAsp--- 210
   ::::::::::::::::::::
Db 380 GTGCTGTCTCCCTCGGTGTTTCACGAGAAAGGACCGCGCTGTGTGGGTTCACGCC 439
QY 211 -----LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGly 226
   ::::::::::::::::::::
Db 440 CGGTGCTTACGATCAGGAGGCGCTTGGAGCTGTTCTGTGATGGCGCGTGTCTCGCGGC 499

```

227 ProGluTrpAsnPheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeu 245  
 500 CCACAGTGAACACACCTCCTCCACTACTGGAAGGAGCGTGGAGGAGGCTGCACAGGCTG 559  
 246 LeuLeuLeuSerTyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAla 265  
 560 TTGTTCTTCGTTACAGAGGAGTGTGATTGAGCCGAGCGGCGGCTGAGAAAGCTGGCC 619  
 266 GluPheMetGlyCysGlyPheThrAspAspGluGluLysGlnGlyLeuValAspGluIle 285  
 620 AATTTTCATGGGCTGTGATTCTCAGAGGAGGAGGAGCGGCGGCTGTGAGCGCATC 679  
 286 ValLysLeuCysSerPheAspAsnLeuLysAsnGlnValAsnLysAsnGlySerSer 305  
 680 GTGAGCTGTGTAGCTTGGGAGCTCAGAGCATGGAGGTAAATAGAATGGAGCACT 739  
 306 TyrAsnSerLysIleAspAsnLysHisPheArgLysGlyGluValArgAspTrpAla 325  
 740 TTGTTGGGG---ATCAAGAACCAAGCTTCTTCAGGAGGAGGAGTGTCTGGGCACTGGAGC 796  
 326 AsnTyrLeuThrSerGluMetIleLysLysLeuGlu 337  
 797 AAGCATATGACGAGGAGATGGCGCAGAGGCTAGAC 832

RESULT 2  
 BZ797778/c 831 bp DNA linear GSS 17-MAR-2003  
 LOCUS BZ797778 ZM 0.6 1.0 KB Zea mays genomic clone ZMMBTA317002,  
 DEFINITION genomic survey sequence.  
 ACCSSION BZ797778  
 VERSION BZ797778.1 GI:29000480  
 KEYWORDS GSS.  
 SOURCE Zea mays  
 ORGANISM Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.  
 1 (bases 1 to 831)  
 Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
 Resnick,A., Fraser,C.M., Yuan,Y., San Miguel,P., Ma,J. and  
 Bennetzen,J.  
 Maize Genomics Consortium  
 Unpublished (2003)  
 Other\_GSSs: FUGZ85TD  
 Contact: Cathy Whitelaw  
 TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
 Tel: 301-838-5843  
 Fax: 301-838-0208  
 Email: whitelaw@tigr.org  
 Seq primer: TR  
 Class: sheared ends.  
 Location/Qualifiers  
 1..831  
 /organism="Zea mays"  
 /mol\_type="genomic DNA"  
 /strain="B73"  
 /db\_xref="taxon:4577"  
 /clone\_lib="ZMMBTA317002"  
 /clone\_lib="ZM 0.6 1.0 KB"  
 /notes="Vector: pCR4-TOPO; Site 1: EcoRI; 0.6-1.0 kb high  
 Cot selected genomic DNA library"

Alignment Scores:  
 Pred. No.: 9,598-46 Length: 831  
 Score: 510.00 Matches: 108  
 Percent Similarity: 57.14% Conservative: 52  
 Best Local Similarity: 38.57% Mismatches: 104  
 Query Match: 24.77% Indels: 16  
 DB: 28 Gaps: 6  
 US-09-854-122-16 (1-391) x BZ797778 (1-831)



```

Db      288  GCACAAAGGACGCTCTGCTCTCTCTGATCTACAAACAAGATCGCGCGATGCTGCG 347
Qy      206  -----AspLysLeuLeuAspLeuGluLysSerValAspLlePheAl 219
Db      348  CCGGAAGTTGCGCTGGAGTCTCCTCGCCACAGTTCGAGGAGGCGTTTCGAGCTTCTG 407
Qy      219  aSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPheThrAsnAlaAl 239
Db      408  CGAGGGGACAGACAGCTCCGACCGCCATGCGGCCACGCCCTTCGAGTACTCGGAGGAG 467
Qy      239  aSerThrAsn---SerAsnLeuLeuLeuLeuSerTyrGluGluMetLeuGluLysProVa 258
Db      468  CCGGAGGAGCGCTGGAGAGTCTCTCTCAGGTACGAGGACATGCTGCAAGACCCAC 527
Qy      258  lGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluGly 278
Db      528  AGGCACAGCCCAACATCTCGCTGCTCATGCGGGTCCCGTTTCTTCGCGAGGAGGA 587
Qy      278  sdGlnGlyLeuValAspGluLeuValLysLeuCysSerPheAspAsnLeuLysAsnGln 298
Db      588  GCGAGGGGTGGTGCAGGAGATCGTGCAGCTCTGTAGCTTTGAGAACTCAAGAGCTCG 647
Qy      298  nValAsnLysAsnGlySerSerTyrAsnSerLysLysLeuAspAsnLysHisPheArg 318
Db      648  GTGACACAGATGGCAGCAGCGCCATGATGGCGCTCAACACGACGCTATCTCAGGAA 707
Qy      318  sGlyGluValArgAspTrpAlaAsnTyrLeuThrSerGluMetLleLysLysLeuGlu 338
Db      708  GCGAGCGGTGGCGACTCGAAATACTACATGACGCGGAGATGGCGGCGCGCTCGAC-- 765
Qy      338  rAlaGlyLysLysLeuAsnGluSerGluLysHisLeuLeuSerAsnLysAsnLeuThr 358
Db      766  -----AGATCGTCGAGGAGCGCTTCAGGTTGAGGCTGACCT----- 804
Qy      358  eAsnLeuAsnAsnTyrCysGluLysSerAsnLysSerLeuCysLeuLeuSerTyrSer 378
Db      805  ----TTGGCATCTCCATGTAGAGGATGGGAGCGAANNCTGTGTATGAAGCATAAATT 860
Qy      378  eValLleAsnAsnPheLleLeuLeuLysLysLys 389
Db      861  T-----TCGTATTGACATCCCAAGAAA 885

```

## RESULT 4

CG290240 975 bp DNA linear GSS 25-AUG-2003  
 LOCUS CG290240 0.7.1.5\_KB Zea mays genomic clone ZM59K24,  
 DEFINITION genomic survey sequence.

ACCESSION CG290240

VERSION CG290240.1 GI:34204454

KEYWORDS GSS.

SOURCE Zea mays

## ORGANISM

Zea mays  
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACCAD  
 clade; Panicoideae; Andropogoneae; Zea.

## REFERENCE

## AUTHORS

Whitehead, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
 Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
 Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.

## TITLE

## JOURNAL

## COMMENT

Unpublished (2002)

Other GSSs: OG2AR72TH

Contact: Cathy Whitelaw

## TIGR

7712 Medical Center Drive, Rockville, MD 20850, USA

Tel: 301-838-5843

Fax: 301-838-0208

Email: whitelaw@tigr.org

Seq primer: TP

Class: sheared ends.

Location/Qualifiers

1..975

/organism="Zea mays"

## FEATURES

## source

```

/mol_type="genomic DNA"
/strain="B73"
/db_xref="taxon:4577"
/clone="ZM59K24"
/clone_lib="ZM_0.7.1.5_KB"
/notes/vector: pBCK-; Site_1: HindII; 0.7-1.5 kb
methylation filtered genomic DNA library"

```

## ORIGIN

Alignment Scores: 1-73e-44 Length: 975  
 Pred. No.: 499.50 Matches: 112  
 Score: 499.50  
 Percent Similarity: 55.67% Conservative: 50  
 Best Local Similarity: 38.49% Mismatches: 111  
 Query Match: 24.26% Indels: 18  
 DB: 29 Gaps: 7

US-09-854-122-16 (1-391) x CG290240 (1-975)

Qy 47 lIleValSerSerLeuProSerAsnAspTyrTrpGlyAsp---ThrMetArgLeuTyrLys 65  
 Db 119 ATCATCCCTTCGCTGCCCTTCGAGACCCAGTGGCCACCGTTCCCTCTCCGACGTCAGCC 178  
 Qy 66 GlyPheTrpGlnMetGlyTyr---LeuValProGlyIleMetAlaPheGluAspAsnPhe 84  
 Db 179 AATTTCGCTGCTGAGGTAGCATTTGAAGGAGGCGTCCCGGGCTTCATTCTCTCTTC 238  
 Qy 85 LysAlaArgGluThrAspLleLeuThrLeuProLysAlaGlyThrTrpThr 104  
 Db 239 GAGCCGAGACCCACCGACGTTCTTCGCAAGTTCCTCCCAAGTTCGACACCTGGCTC 298  
 Qy 105 LysAlaLeuThrPheAlaLleLeuThrArgAspValAsnHisProSerSerProThrHis 124  
 Db 299 AAGCCCTTCGCTTCGAGCGCTGAAGCGTTCCACGATCCGCGCTTCGACGACGACCAT 358  
 Qy 125 ProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGly 144  
 Db 359 CCGCTGCGCATTCGCAACCTCATGACTGTGTGAGTTCCTCGAACTTGATTTCAACAC 418  
 Qy 145 ArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPheAlaGly 164  
 Db 419 CAGAGGATGAGTTCGAGGCGCTC-----CCGTCCCGCGTGTGCTGGCCACG 466  
 Qy 165 HisIleProTyrSerLeuLeuProAlaSerValLeu-----LysSerGlyThr 180  
 Db 467 CACCTCCCTTACTCCCTGCTGCGCGGAGCATCCCGGGACGACGAGAGGCGCTCGGGTGC 526  
 Qy 181 LysIleLeuAsnLysSerArgAsnArgLysSerThrPheValSerPheTrpLysPheGly 200  
 Db 527 CGGATCTCTACGTTTCGCGGGAACCCAGGACGTGCTGTCTCTCTCTGCTGTTCAAG 586  
 Qy 201 AsnLeuIleAsnProAspLysLeuLeuAsp-----LeuGluLysSerVal 215  
 Db 587 AGAAAGGACGCGCGGCTTCTGGGTGTGACCGCGGTTCACGATCCAGGAGGCGCTTG 646  
 Qy 216 AspLlePheAlaSerGlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPhe 235  
 Db 647 GAGCTGTCTGTGATGCCCTGTCTCGCGGCGCCACGATGGAACACACGCTCTCCACTAC 706  
 Qy 236 ---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeu 254  
 Db 707 TGAAGGAGGAGCGTGAAGGCGCTGACAGGCGTGTGTTCTCTCTACGAGGAGGCTG 766  
 Qy 255 GluLysProValGluAsnValLysLeuAlaGluPheMetGlyCysGlyPheThrAsp 274  
 Db 767 ATTGAGCCCGAGGCGGACGTCAGAAAGCTGGCAATTCATGGCGCTGTGATTTCTCAGAG 826  
 Qy 275 AspGluGluLysGlnGlyIleValAspGluIleValLysLeuCysSerPheAspAsnLeu 294  
 Db 827 GAGAGGAGGAGCGCGGCGTGGTGGAGCCATCGTGGAGCTGTGTAGCTTGTGGGAGACTG 886  
 Qy 295 LysAsnGlnValAsnLysAsnGlySerSerTyrAsnSerLysLysLysLysAspAsnLysHis 314

```

Db      887 AGAGACATGGAGGTAAATAGAAATGAGACACTTTGTTGGG---ATCAAGAACGAAAGC 943
QY      315 Phe-PheArgLysGlyGluValAla9AspTTP 324
Db      944 TTCTTTCAGGAAGGAGTGTGTTGGGACTGG 974

RESULT 5
LOCUS   CC698091/c
DEFINITION OGBP20TV ZM_0.7_1.5_KB Zea mays genomic clone ZMMBMA0347C16,
genomic survey sequence.
ACCESSION CC698091
VERSION   CC698091.1 GI:32102867
KEYWORDS  GSS.
SOURCE    Zea mays
ORGANISM  Zea mays
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD
clade; Panicoideae; Andropogoneae; Zea.
1 (bases 1 to 990)
Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,
Citek, R.W., Nurnberg, A., Robbins, D. and Lakey, N.
Consortium for Maize Genomics
Unpublished (2002)
Other GSSs: OGBP20TH
Contact: Cathy Whitelaw
TIGR
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-838-5843
Fax: 301-838-0208
Email: whitelaw@tigr.org
Seq primer: TP
Class: sheared ends.
FEATURES
    source
        location/Qualifiers
            1..990
                /organism="Zea mays"
                /mol_type="genomic DNA"
                /strain="B73"
                /db_xref="caxon:4577"
                /clone="ZMMBMA0347C16"
                /clone_lib="ZM 0.7 1.5 KB"
                /note="Vector: pBCKS-; Site 1: HincII; 0.7-1.5 kb
                methylation filtered genomic DNA library"

ORIGIN
Alignment Scores:
Pred. No.:      6.29e-43      Length:      990
Score:          485.50      Matches:    121
Percent Similarity: 51.59%      Conservative: 41
Best Local Similarity: 38.54%      Mismatches: 127
Query Match:     23.58%      Indels:     25
DB:              29          Gaps:         9

US-09-854-122-16 (1-391) x CC698091 (1-990)
QY      61 MetArgLeuTyrLysGlyPheTrpGlnMetGlyTyrLeuValProGlyIleMetAlaPhe 80
Db      998 CTGCGCCCTGTACAGAGCTTCTGGTCGCGGAGCACTGGGTGCCGGGACCATCGTGTTC 929
QY      81 GluAspAsnPheLysAlaArgGluThrAspIlelleLeuThrLeuProLysAlaGly 100
Db      928 CAGCGCCGCTTTCACCGCGCCCGCCGACGACGTGATCTCTCGGAGCTACCCCAAGTCGCGC 869
QY      101 ThrTrpThrLysAlaLeuThrPheAlaIle-LeuThrArgAspValAsnHisProSe 120
Db      868 ACCAGTGGCTGAAGCGCTGGCTTCGCCACGGCGCGGTACGGGTACCGCGCGG 809
QY      120 rSerProThrHisProLeuPheAsnProHisSerCysValGlnAsnLeuGluTy 140
Db      808 TGGCGCCAGGACCCCGCTCCGACGGCTCAACCCGACGACTCGCTCCGGTTCATCGACGA 749
QY      140 rleuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProAr 160

```

```

Db      748 GATCTTCGCCAGCGGGAGGAGCG-----AAGCTCGAGGCCCTC---CCGTGCGCCGAG 698
QY      160 glenPheAlaGlyHisIleProTyrSerLeuLeuProAlaSerVal-----LeuLysSe 178
Db      697 GCTGATGAACACGACCTGCTAGCGCTGCTCCGCGCCCGCTCACCGCCACCGCCAC 638
QY      178 rGlyThrLysIlelleAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLy 198
Db      637 CGGTTCAGAGGTTCGCTAGCTGTGCAGGAGACCCCAAGGACATGGTGGTCTGCTCTGGCA 578
QY      198 sPheGlyAsnLeuIleAsnProAspLysLeuLeuAspLeuGluLysSerValAspIlePh 218
Db      577 CTTCCTCCGCGCGCCAAACGCGACCTCTTGTTC-----GCCGACACGTT 533
QY      218 eAlaSer-----GlyIleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPh 235
Db      532 CGAGTCCGTCTCGACGCGACCGTGTGTGGCCGCGGTCTGGGACCACGCTCTCTCCCTA 473
QY      235 eThrAsnAlaIaIa----SerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetIle 254
Db      472 CTGGCGTGTGAGCGTGGCGGCCCGGACAGGCGCGCTCTTCTCAGGTACGAGGACATGCT 413
QY      254 uGluLysProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAs 274
Db      412 GCGGACCCCGCGCGCACGTCGCGGGCTGCCGAGTTTCATGGCGCGCCCTTCTCGGC 353
QY      274 pAspGluLysGlnGlyIleValaspGluIleValLysLeuLysCysSerPheAspAsnLe 294
Db      352 CGCGGAGGAGCGCGCGCGAGCTGCGCGCCCTCGTGAGCTCTCGAGCTTCGACGTTCCAGCAGAT 293
QY      294 uLysAsnGlnGlnValAsnLysAsnGlySerSerTyrAsnSerLys-----IleAs 311
Db      292 GAAGGCTGGAGGTGAACAGGCTGTGAGCGGACCGCGGGGAGGTACCGGCCCATGCGC 233
QY      311 pAsnLysHisPhePheArgLysGlyGluValArgAspTrpAlaAsnTyrLeuThrSerGl 331
Db      232 GCGCGACGCGCTTCTTCGGAAGGCTGTCGCGGAGCTGGCGGAAACCATGACGCGCGA 173
QY      331 uMetIleLysLysLeuGluThrAlaGlyLysIleAsnGlu-SerGluLysHisLeuLeuS 351
Db      172 GATGGCGCGCGCTGGAC-----GGGATCTTCGCTGAAAGTCCCAAG----- 129
QY      351 exAsnLysAsnLeuThrAsnPheAsnLeuAsnAsnTyr 363
Db      128 --GGACGGGGCTTATCTTCTCATATGCAAAATGTGTAT 93

RESULT 6
CF833911/c
LOCUS   CF833911
DEFINITION UCRS02_04105 f Ruby Orange Ovary at Anthesis cDNA Library Citrus
sinensis cDNA clone CS_REA04105, mRNA sequence.
ACCESSION CF833911
VERSION   CF833911.1 GI:38036870
KEYWORDS  EST.
SOURCE    Citrus sinensis
ORGANISM  Citrus sinensis
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots;
Rosids; eurosida II; Sapindales; Rutaceae; Citrus.
1 (bases 1 to 859)
Close, T.J., Roose, M.L., Federici, C.F., Fenton, R.D., Wanamaker, S.,
Kim, H.R., Kudrna, D., Wing, R. and Yu, Y.
Development of EST Resources and New Genetic Markers for California
Citrus - Ruby Orange Ovary at Anthesis
Unpublished (2003)
Contact: Timothy Close
Department of Botany & Plant Sciences, University of California
Riverside, CA, 92521-0124
Tel: 9097873318
Fax: 9097874437
Email: timothy.close@ucr.edu
Seq primer: T7.

```

FEATURES  
source

Location/Qualifiers  
1. 859  
/organism="Citrus sinensis"  
/mol\_type="mRNA"  
/cultivar="Ruby"  
/db\_xref="taxon:2711"  
/clone="CS\_REA04105"  
/tissue\_type="Ovary"  
/dev\_stage="20 Year old trees"  
/lab\_host="E. coli TJC121"  
/clone\_lib="Ruby Orange Ovary at Anthesis cDNA Library"  
/note="Vector: Lambda Uni-ZAP XR, excised phagemid;  
Site 1: EcoRI; Site 2: XhoI; Two trees with  
open-pollinated flowers in a mixed planting at UC  
Riverside Agricultural Operations were the source of  
tissue. Ovaries at anthesis were collected by Federici in  
March 2003. Tissues were snap frozen and then stored at  
-80C until further processing. Renton (Close lab) purified  
RNA by the phenol method described in J. Japanese Soc.  
Hort. Sci. 1996. 64 (4): 809-814, purified poly(A) mRNA  
using a PolyATrack mRNA Isolation System IV (Promega),  
produced a primary cDNA library using a lambda ZAP XR cDNA  
Synthesis Kit (Stratagene), then mass-excised 450,000 pfu  
from the primary library to produce a phagemid population.  
Phagemids were plated, plasmid DNA purified, cDNA clones  
archived, and DNA sequences determined bi-directionally  
using an ABI3730 at the Arizona Genomics Institute,  
University of Arizona (Kim, Kudrna, Wing, Yu).  
Chromatogram files were downloaded by FTP to UC Riverside  
(by Close), then processed at UC Riverside (by Wanmaker,  
Close lab) using the HarVest pipeline  
(http://harvest.ucr.edu) to remove vector and cloning  
oligo sequences and various contaminants, and to trim to a  
high quality region. Sequences that retained a phred 17  
region of at least 100 bases were deposited to GenBank."

## ORIGIN

Alignment Scores:  
Pred. No.: 1,28e-42 Length: 859  
Score: 482.00 Matches: 125  
Percent Similarity: 55.41% Conservative: 44  
Best Local Similarity: 40.98% Mismatches: 96  
Query Match: 23.41% Indels: 40  
DB: 14 Gaps: 14

US-09-854-122-16 (1-391) x CF933911 (1-859)

QY 107 LeuThrPheAlaIleLeuThrArgAspValAsnHisProSerSerPro---ThrHisPro 125  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 859 ATAACGTTTGCCATAGTCACCGT---TTGCGTATGTTGATAATCCCTGTACACCGCT 803  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 126 LeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyr---LeuTyrMetGly 144  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 802 TTGCTTACCACCAATCTCTCATGAGCTGTACCGCTTCATTGAGCTTAATATATGTT--- 746  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 145 ArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPheAlaGly 164  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 745 ---GATAATCAGGTTCTGTGACCTCACACCTTGACC---TCTCCAGGCTCTTTCACCC 692  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 165 HisIleProTyrSerLeuLeuProAlaSerValLeuLysSerGlyThr-----LysIle 182  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 691 CATTTGCCCTTTGTCTCATTTGCTTGATCTGTTAAAGATTTCATCAGCTTGCAACTT 632  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 183 IleAsnLysSerArgAsnArgLysSerThrPheValSerPheTrpLysPheGlyAsnLeu 202  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 631 GTGTATTGTGTAGGAACCCCTAAGGACATTTTGTGCTGCTCTGGCACCTTTACCAACAG 572  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 203 IleAsnProAspLys-----LeuLeuAspLeuGluLysSerValAspIlePheAlaSer 220  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 571 CTGAGACCCGAGGAAGGGGTACAAATCTCTTGAAGAATCTTTGATAGTTTTCAGG 512  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 221 GlyIleSerPheCysGlyProGluTyrAsnPheGlnAlaGluPheThrAsnAlaSer 240  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 511 GGAGTGAGTTTGTATGGGCCCTTTTGGGATCATATTTTGGGCTACTGGAGCAAGCATG 452  
QY 241 ThrAsn---SerAsnLeuLeuLeuLeuSerTyrGluGluMetLeuLysProValGlu 259  
Db 451 GAAAAACCTGATAGGTTTCTTCTTAAAGTATGAAGAACTGAAGAAAGAGCCAGCTCTT 392  
QY 260 AsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGluLysGln 279  
Db 391 CGTTTGAAGATGCTGCTGAGTCTTGGGATGTCCATTTTGGCCAGCAGAGAGCGAAT 332  
QY 280 GlyIleValAspGluLeuValLysLeuCysSerPheAspAsnLeuLysAsnGlnVal 299  
Db 331 GGCTGTGGTATGATATCTTGAAGTTATGTAGCTTTGATTAATCTTGACCACTTGGAGGTC 272  
QY 300 AsnLysAsn-----GlySerSerTyrAsnSerLysLysLeuAspAsnLysHis 314  
Db 271 AACAGAAACCGGAAATGTCTCCGCGAAGGGTACAAACACC----- 230  
QY 315 PhePheArgLysGlyGluValArgAspTyrAlaAsnTyrLeuThrSerGluMetIleLys 334  
Db 229 TTTTTCGCGGAGGTGTGGTGGAGATTTGGATGAATTTACTGACCCCGAAATGGTAGGA 170  
QY 335 LysLeuGluThrAlaGlyLysIleAsnGluSerGluLysHisLeuLeuSerAsnLys--- 353  
Db 169 CGTTTGGAC-----CAATCATTAACAAAAATTCATGCTGCTTTGGATTAAGATTC 119  
QY 354 -----AsnLeuThrAsnPheAsnLeuAsnAsnTyrCysGluLysSer 367  
Db 118 TAGACTCGTAGACTGGTAAATGTTTCATGTTTATGTGTGATGCTGTTGTCGTCCTTT 59  
QY 368 AsnIleSerLeuCysLeuLeu-----SerTyrSerPheValIleAsnAsnPheIleLeu 385  
Db 58 ACATAAAGTTGGTGTGTTAGATTAACTACATCGTT-----CTC 17  
QY 386 LeuLysLysLysLys 390  
Db 16 TTTAAAAAATAAAAAA 2

RESULT 7  
CC691394/c 870 bp DNA linear GSS 19-JUN-2003  
LOCUS OGU9595TV ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0425021,  
DEFINITION genomic survey sequence.  
ACCESSION CC691394  
VERSION CC691394.1 GI:32096170  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
Eukaryote; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; PACAD  
clade; Panicoideae; Andropogoneae; Zea.  
REFERENCE 1 (bases 1 to 870)  
AUTHORS Whitelaw, C.A., Quackenbush, J., Van Aken, S., Utterback, T.,  
Resnick, A., Fraser, C.M., Budiman, M.A., Bedell, J.A., Rohlfing, T.,  
Citek, R.W., Nunberg, A., Robbins, D. and Lakey, N.  
TITLE Consortium for Maize Genomics  
JOURNAL Unpublished (2002)  
COMMENT Other GSSs: OGU9595TH  
Contact: Cathy Whitelaw  
TIGR 9712 Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: Tg  
Class: sheared ends.  
Location/Qualifiers  
1. .870  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0425021"





```
Db 641 GCCAGGCTGCTCGGAGCACCTGCGCCCTCTCCCTGCTCCGCGCGCGTGGCCACCTC 582
Qy 178 rGlyThrLysIleIleAsnIleSerArgAsnArgLysSerThrPheValSerPheTyrPly 198
Db 581 CGGCTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 522
Qy 198 sPheGlyAsnLeuIleAsnProAspLysLeu---LeuAspLeuGluLysSerValAspI 217
Db 521 CTTTCGTAAGGCGATCGCGGGGCGAGGTCCCGCGTGGAGCTGGACGCGCGCTGGCCAT 462
Qy 217 ePheAlaSerGlyIleSerPheCysGlyProGluTyrAsnPheGlnAlaGluPhe---Th 236
Db 461 GTTCTCGAGGGGCTGCGCTTTCGCGCCGCTGCGGAGCTACTACCTGAGTATGGAA 402
Qy 236 rAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLys 256
Db 401 GGAGAGCTGGCGAGGCTGAGCGGTCTCTTTCTCAGGTACGAGGAGATGCTGGCGGA 342
Qy 256 sProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAsp 276
Db 341 CCGGCTGCGTGGCTGAGGAGCTGCGCGCTCTTCGCGCTGCGCTGCGCTGCGCTGCG 282
Qy 276 uGluLysGlnGlyIleValAspGluIleValLysLeuLysSerPheAspAsnLeuLysAs 296
Db 281 AGAGGCGGAGGCTGCGGAGGAGATGTGAGGCTGTCAGCTTCGAGATGCTGAGCGG 222
Qy 296 nGlnInValAsnLysAsnGlySer-----SerTyrAsnSerLysIleAspAsnLys 313
Db 221 CTTTGAAGAAGATCGGACCGGAGATCTTGATTCGGGAGACACACCGGTGCGGAATC 162
Qy 313 sHsPhePheArgLysGlyGluValArgAspTyrPalaAsnTyrLeuThrSerGluMetI 333
Db 161 TACCTTCTTCAAGAGGCGAAGTTGGGACTGGGAGACCATCTTGACAAAGAGATGGG 102
Qy 333 eLysLysLeuGlu 337
Db 101 CAAGAGGCTCGAC 89
```

```
RESULT 9
CG919663/c
LOCUS MBEBT65TRB mth2 Medicago truncatula genomic clone 22L10, genomic
DEFINITION survey sequence.
ACCESSION CG919663
VERSION CG919663.1 GI:39779346
KEYWORDS GSS.
SOURCE Medicago truncatula (barrel medic)
```

```
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
1 (bases 1 to 934)
Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.F.
```

```
REFERENCE
AUTHORS Sequencing of BAC ends from Medicago truncatula
TITLE Unpublished (2003)
JOURNAL Other GSSs: MBEBT65TF
COMMENT Contact: Chris Town
TIGR
```

```
9712 Medical Center Drive, Rockville, MD 20850, USA.
Tel: 301-838-1523
Fax: 301-838-0208
Email: cdtown@tigr.org
Seq primer: CAGGAACAGCTATGACC
Class: BAC ends.
```

```
FEATURES
source
1..934
Location/Qualifiers
1..934
/organism="Medicago truncatula"
/mol_type="genomic DNA"
/cultivar="genotype Al7"
/db_xref="taxon:3880"
/clone="22L10"
```

```
ORIGIN
Alignment Scores:
Pred. No.: 3,486-41 Length: 934
Score: 469.50 Matches: 99
Percent Similarity: 59.70% Conservative: 58
Best Local Similarity: 37.64% Mismatches: 99
Query Match: 22.80% Indels: 7
DB: 29 Gaps: 5
```

US-09-854-122-16 (1-391) x CG919663 (1-934)

```
Qy 34 LysGluGluAspSerLysMetTyrLysArgTyrArgGluIleValSerSerLeuProSer 53
Db 780 AAAGAGATGAC-----TTAGGCAAGAAATGCAAGGACTTGATACAAACCTTGCCATTA 727
Qy 54 AsnAspTyrTyrPlyGlyAsp---ThrMetArgLeuTyrLysGlyPheTyrGlnMetGlyTyr 72
Db 726 GTAGAAGGTGGATCGATCCACCTTTTCATGAATATCAAGATTTTGGTTGCTCCCAAG 667
Qy 73 LeuValProGlyIleMetAlaPheGluAspAsnPhelLysAlaArgGluThrAspIleIle 92
Db 666 ATTCTCCAAAGTGCTTTATCTTTTCAAAAGCACTTTCAAGCCATTGTAGACATATTATC 607
Qy 93 LeuThrThrLeuProLysAlaGlyThrTyrLysAlaLeuThrLysAlaLeuThrPheAlaIleLeu 112
Db 606 TTAGTTACAATCCCAATCAGTACCCTGGCTCAAGCATTTGCTTTGCTTTGATA 547
Qy 113 ThrArgAspValAsnHisProSerSerProThrHisProLeuLeuPhePheAsnProHis 132
Db 546 AACCGTAACAAATATCCAAATATTCATAAAATCATCTTTTACTCCTTCCAAACCCCTCAT 487
Qy 133 SerCysValGlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeu 152
Db 486 GTTCTGTCTCTTCATGAGATTAACCTTACTATGAACTGATATGTCTCTGAATC 427
Qy 153 AspMetLeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuPro 172
Db 456 AATTCACCTATCC---CCTCCAAAGAAATCTTCCACACACAAATCCTTATGTTTGTGCCA 370
Qy 173 AlaSerValLeuLysSerGlyThrLysIleLeuAsnIleSerArgAsnArgLysSerThr 192
Db 369 AAATCTTTGAAGAGATCCCACTGTAAGTGTGTATCTTTGTAGAGATCTCTTAAGACACT 310
Qy 193 PheValSerPheTyrLysPheGlyAsnLeuIleAsnProAspLys-----LeuLeuAsp 210
Db 309 TTTGTTTTTCATTTGGCAATTTCTCAACAAGCTAAAGAGAACAAAGTAGAGGACACTTCCA 250
Qy 211 LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTyrAsn 230
Db 249 TTAGAAGAGGCTTTTGAGAGTTTGTAGAGGAAACAACCTTCTTATGGACCCCTTTTGGAT 190
Qy 231 PheGlnAlaGluPhe---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSer 249
Db 189 CATGTATTAGGATACCTGGAGGAAAGCTTGGAGAGACCGGAGAGAGTATGTTTGGAAA 130
Qy 250 TyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGly 269
Db 129 TATGAAGAAATGAAATGAAACCTTAATTTGTTTGAAGAGAGATTCGTAAGTTTGGGG 70
Qy 270 CysGlyPheThrAspAspGluLysGlnGlyIleValAspGluIleValLysLeuCys 289
Db 69 TGTCCATTTCCGAGAGAGAAATCCCAATGCTGTGTGTATGATATATTACATTGTGT 10
Qy 290 SerPheAsp 292
Db 9 AGTTTGTAG 1
```

RESULT 10  
BF065941

```

LOCUS       BF065941                950 bp      mRNA      linear      EST 23-OCT-2001
DEFINITION  HV_CEB0014G16f Hordeum vulgare seedling green leaf EST library
CDNA clone  HV_CEB0014G16f, mRNA sequence.
ACCESSION   BF065941
VERSION     BF065941.1  GI:10842580
KEYWORDS    EST.
SOURCE      Hordeum vulgare subsp. vulgare
ORGANISM    Hordeum vulgare subsp. vulgare
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
Pooidae; Triticeae; Hordeum.
1. (bases 1 to 950)
   Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Wei, F., Begum, D.,
   Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J.,
   Choi, D.W., Xenton, R.D., Oates, R. and Main, D.
   Development of a genetically and physically anchored EST resource
   for barley genomics: Blumeria infectionally incompatible (Mla6) seedling
   leaf cDNA library.
   Unpublished (2001)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Total hg bases = 574
Seq primer: AATTACCCCTCATTAAAGG
High quality sequence stop: 644.
Location/Qualifiers
1. 950
   /organism="Hordeum vulgare subsp. vulgare"
   /mol_type="mRNA"
   /cultivar="C116151 (Mla6)"
   /sub_species="vulgare"
   /db_xref="taxon:115509"
   /clone="HV_CEB0014G16f"
   /tissue_type="seedling green leaf"
   /lab_host="SOLR"
   /clone_lib="Hordeum vulgare seedling green leaf EST
   library HVCDNA0005 (Blumeria challenged)"
   /note="vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
   C.I. 16151 (Mla6) plants were greenhouse grown in the R
   green lab at Iowa State University, Ames, IA; 7 day old
   green seedlings were challenged with isolate 5874
   (AvrMla6) of Blumeria graminis f. sp. hordei, and leaves
   were harvested 20 and 24 hr post-inoculation and snap
   frozen; uninoculated leaves were harvested 20 hr
   post-inoculation (Wei, Wise). In the TJ Close lab at the
   University of California, Riverside, total RNA was
   prepared from each sample pool, equal quantities of all
   three RNA pools were combined, poly(A) RNA was purified
   from the mixture, one primary unamplified cDNA library was
   made, and 1 million pfu were in vivo excised to give
   pBlueScript SK(-) cDNA phagemids (Choi, Close). Phagemids
   were plated and picked at the Clemson University Genomics
   Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing).
   Plasmid DNA preparations, DNA sequencing and sequence
   analysis were performed at CUGI (Wing, Yu, Frisch, Henry,
   Simmons, Oates, Rambo, Main). The sequence has been
   trimmed to remove vector sequence and contains a minimum
   of 100 bases of phred value 20 or above. For more details
   on library preparation and sequence analysis see
   http://www.genome.clemson.edu/projects/barley. To order
   this clone see http://www.genome.clemson.edu/orders/Also
   see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
   Genetically and physically anchored EST resources for
   barley genomics. Barley Genetics Newsletter 31:29-30.
   (http://wheat.nw.edu/genetics/bam/31/cover.html)
FEATURES             (1)
Source              1..950
                     /organism="Hordeum vulgare subsp. vulgare"
                     /mol_type="mRNA"
                     /cultivar="C116151 (Mla6)"
                     /sub_species="vulgare"
                     /db_xref="taxon:115509"
                     /clone="HV_CEB0014G16f"
                     /tissue_type="seedling green leaf"
                     /lab_host="SOLR"
                     /clone_lib="Hordeum vulgare seedling green leaf EST
                     library HVCDNA0005 (Blumeria challenged)"
                     /note="vector: lambdaZAP; Site 1: EcoRI; Site 2: XhoI;
                     C.I. 16151 (Mla6) plants were greenhouse grown in the R
                     green lab at Iowa State University, Ames, IA; 7 day old
                     green seedlings were challenged with isolate 5874
                     (AvrMla6) of Blumeria graminis f. sp. hordei, and leaves
                     were harvested 20 and 24 hr post-inoculation and snap
                     frozen; uninoculated leaves were harvested 20 hr
                     post-inoculation (Wei, Wise). In the TJ Close lab at the
                     University of California, Riverside, total RNA was
                     prepared from each sample pool, equal quantities of all
                     three RNA pools were combined, poly(A) RNA was purified
                     from the mixture, one primary unamplified cDNA library was
                     made, and 1 million pfu were in vivo excised to give
                     pBlueScript SK(-) cDNA phagemids (Choi, Close). Phagemids
                     were plated and picked at the Clemson University Genomics
                     Institute (CUGI) (Begum, Palmer, Frisch, Atkins and Wing).
                     Plasmid DNA preparations, DNA sequencing and sequence
                     analysis were performed at CUGI (Wing, Yu, Frisch, Henry,
                     Simmons, Oates, Rambo, Main). The sequence has been
                     trimmed to remove vector sequence and contains a minimum
                     of 100 bases of phred value 20 or above. For more details
                     on library preparation and sequence analysis see
                     http://www.genome.clemson.edu/projects/barley. To order
                     this clone see http://www.genome.clemson.edu/orders/Also
                     see Close TJ, Wing R, Kleinhofs A, Wise R (2001)
                     Genetically and physically anchored EST resources for
                     barley genomics. Barley Genetics Newsletter 31:29-30.
                     (http://wheat.nw.edu/genetics/bam/31/cover.html)

```

sequence.  
CB330744.1 GI:28879973  
EST.  
Ipomoea batatas (sweet potato)  
Ipomoea batatas  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;  
asterids; lamids; Solanales; Convolvulaceae; Ipomoea.  
1 (bases 1 to 1118)  
Jankowicz, J., Berenyi, M. and Burg, K.  
Identification of adaptation specific differences in the mRNA  
expression profile of drought tolerant and sensitive sweet potato  
Unpublished (2004)  
Contact: Berenyi M.  
Department of Biotechnology  
ARC Seibersdorf Research GmbH  
A2444 Seibersdorf, Austria  
Tel: +43 50550 3524  
Fax: +43 50550 3444  
Email: Maria.Berenyi@arcs.ac.at  
PCR Primers  
FORWARD: lambda Triplex 5' Screening Amplimer  
BACKWARD: lambda Triplex 3' Screening Amplimer  
Insert Length: 1118 Std Error: 0.00  
Seq primer: lambda Triplex Sequencing Primers  
POLYA=yes.  
Location/Qualifiers  
1..1118  
/organism="Ipomoea batatas"  
/mol\_type="mRNA"  
/db\_xref="taxon:4120"  
/tissue\_type="Leaf"  
/lab\_host="Lambda"  
/clone\_lib="Sweetpotato cDNA library"  
/notes="Vector: Triplex2; Site 1: Sfi A; Site 2: Sfi B;  
Ipomoea batata (Lam) L.; Smart cDNA library Construction  
kit; Clontech: PT3000-1"

ORIGIN  
Alignment Scores:  
Pred. No.: 9,456-41 Length: 1118  
Score: 466.50 Matches: 126  
Percent Similarity: 49.22% Conservative: 63  
Best Local Similarity: 32.81% Mismatches: 151  
Query Match: 22.66% Indels: 45  
DB: 14 Gaps: 10  
US-09-854-122-16 (1-391) x CB330744 (1-1118)  
QY 38 SerLysMetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAspTyrTrp 57  
Db 1 AGTCATCAATATGATATGAGACAGACAGCATATATCTCA-GAGGAAGAAGAGTGG 59  
QY 58 GlyAspThr---MetArgLeuTyrLysGlyPheTyrGlnMetGlyTyrLeuValProGly 76  
Db 60 GGAGACAAACATTTAGACGACATCAATGGTCTCTGCTTTTCTAGCCCTTTT 119  
QY 77 IleMetAlaPheGluAspAsnPheLysAlaArgGluThrAspIleLeuThrThrLeu 96  
Db 120 ACTCGCGAGTGTAGCGGAGTTTAACCCACGTCGCCAGGACGATATTTTGGTTCGTT 179  
QY 97 ProLysAlaGlyThrThrTrpThrLysAlaLeuThrPheAlaIleLeuThrArgAspVal 116  
Db 180 CCCAAACCGGCTCCACATGGCTCAATCTCTCTTACTCCATCATCAATCGC----- 233  
QY 117 AsnHisProSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGln 136  
Db 234 -----TCTCTCTTCATCTCTTGTCAACACMACCCACCCACGACCTCGTGCCT 281  
QY 137 AsnLeuGlu---TyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMetLeu.155  
Db 282 TTCCTAGAAGTTCGAAGTCTATGACAGCAAGATCATCACTCACTGTGTCCGAGGAT 341



```
Db 33 ATCATCCCTGCTGCTCCCTCGAGACCAAGGTGGCCAGCTTCCCTCTCCGACGGTACGCC 92
Qy 66 GlyPheTyrGlnMetGlyTyr---LeuValProGlyIleMetAlaPheGluAspAsnPhe 84
Db 93 AATTTTGGCTGCTGAGTAACTATTGAAGGAAGCGTCCCGGGCTTCATTCTGCTCTC 152
Qy 85 LysAlaArgGluThrAspIlelelelelelelelelelelelelelelelelelelele 104
Db 153 GAGCGAGACCCACCGAGCTGCTCTCGAAGCTTCCCAAGTCTGGCACCACCTGGCTC 212
Qy 105 LysAlaLeuThrPheAlaIlelelelelelelelelelelelelelelelelelelele 124
Db 213 AAGGCCCTCGCTCGAGCGTGAAGCGTTCACCGCATCCCGCTTCGACGACGACCAT 272
Qy 125 ProLeuLeuPhePheAsnProHisSerCysValGlnAsnLeuGluTyrLeuTyrMetGly 144
Db 273 CCGCTGCGCATTCGACACCTCATGACTGTGTGAGGTCTCTCGAA-----CTTGGT 323
Qy 145 ArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSerProArgLeuPheAlaGly 164
Db 324 TTCACACAGACAGATGATGCTCGAGCGCTC---CGTCCCGGGTGTCTGCGCCACG 380
Qy 165 HisIleProTyrSerLeuLeuProAlaSerValLeu-----LysSerGlyThr 180
Db 381 CACCTCCCTACTCCCTGCTGCGCGCAGCATCACCGGGACGAGAGACCTCGGGGTGC 440
Qy 181 LysIlelelelelelelelelelelelelelelelelelelelelelelelelelele 200
Db 441 CGCATGCTGTAGCTGCGCGGAGCCCAAGGACATGCTGTCTCTCTGCTGCTGCTGCT 500
Qy 201 AsnLeuLeuAsnProAspLysLeuLeuAsp-----LeuGluLysSerVal 215
Db 501 AGGAAGGACGCGCCCTTGTGGGTGTGACCGCGCTGTTCCAGCATCCAGAGGCGCTTG 560
Qy 216 AspIlePheAlaSerGlyIleSerPheCysGlyProGluTyrAsnPheGlnAlaGluPhe 235
Db 561 GAGCTGTTCTGTCATGCGCTGCTCCCGCGGCCACAGTGAACACCGCTCTCCAGTAC 620
Qy 236 ---ThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuLeuSerTyrGluGluMetLeu 254
Db 621 TGGAGGAGAGCGCTGAGGAGCGCTGACAGGGTGTGTTCTCTGTTTACGAGGAGGTG 680
Qy 255 GluLysProValGluAsnValLysLeuAlaGluPheMetGlyCysGlyPheThrAsp 274
Db 681 ATTGAGCCCGAGCGCGCGGTGGTGGAGCACCATCGTGGAGTGTGTAGCTGGCGGAGCTG 740
Qy 275 AspGluGluLysGlnGlyIleValAspGluIleValLysLeuLysSerPheAspAsnLeu 294
Db 741 GAAGAGAGACACCGGGGTGGTGGAGCACCATCGTGGAGTGTGTAGCTGGCGGAGCTG 800
Qy 295 LysAsn 296
Db 801 AGGGAC 806

RESULT 14
CG950535 855 bp DNA linear GSS 15-DEC-2003
LOCUS MBG003TF mth2 Medicago truncatula genomic clone 51B5, genomic
DEFINITION survey sequence.
ACCESSION CG950535
VERSION CG950535.1 GI:39861114
KEYWORDS Medicago
SOURCE Medicago truncatula (barrel medic)
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots;
rosids; eurosids I; Fabales; Fabaceae; Papilionoideae; Trifolieae;
Medicago.
REFERENCE 1 (bases 1 to 855)
AUTHORS Town, C.D., Shetty, J., Koo, H. and Feldblyum, T.F.
TITLE Sequencing of BAC ends from Medicago truncatula
JOURNAL Unpublished (2003)
COMMENT Other_GSSs: MBG003TR
```

Contact: Chris Town  
TIGR  
9712 Medical Center Drive, Rockville, MD 20850, USA.  
Tel: 301-838-3523  
Fax: 301-838-0208  
Email: cdtown@tigr.org  
Seq primer: TGTAAACACGACGCGCCAGT  
Class: BAC ends.  
Location/Qualifiers  
1. 855  
/organism="Medicago truncatula"  
/mol\_type="genomic DNA"  
/cultivar="genotype A17"  
/db\_xref="taxon:3880"  
/clone\_lib="51B5"  
/clone\_lib="mth2"  
/notes="Vector: pBelBAC11; Site 1: HindIII; Site 2:  
HindIII; Cook, D.R. and Kim, D.J., unpublished"

FEATURES  
source

## ORIGIN

Alignment Scores:  
Pred. No.: 4,49e-37 Length: 855  
Score: 432.00 Matches: 101  
Percent Similarity: 58.49% Conservative: 54  
Best Local Similarity: 38.11% Mismatches: 102  
Query Match: 20.98% Indels: 9  
DB: 29 Gaps: 5  
US-09-854-122-16 (1-391) x CG950535 (1-855)

Qy 36 GluAspSerLysMetTyrLysArgTyrArgGluIleValSerSerLeuProSerAsnAsp 55  
Db 66 CAGATGATGACTTATGACCAAGAAATGCAAGAGTTGATACCCACCTTGCCTCAGAAAA 125  
Qy 56 TyrTyrGlyAspThrMetArgLeu---TyrLysGlyPheTyrGlnMetGlyTyrLeuVal 74  
Db 126 GCCTGGATTTTAAACCCACACCTATCAATATCAAGATTTTGGATTGCCAAAAAATTT 185  
Qy 75 ProGlyIleMetAlaPheGluAspAsnPheLysAlaArgGluThrAspIlelelelele 94  
Db 186 CAGGCTGTTTATCTTCTGCAAGACCTTTCAGAGCTTATGATTCAGATATCATCTTAGT 245  
Qy 95 ThrLeuProLysAlaGlyThrThrTyrThrLysAlaLeuThrPheAlaIlelelelele 114  
Db 246 ACCGCTCCCAATACAGTACCCTTGGCTCAAAGCATTTGCTTGTCTTAAACCGC 305  
Qy 115 AspValAsnHisProSerSerProThrHisProleuLeuPhePheAsnProHisSerCys 134  
Db 306 AACAAATATCCAAATTATTCATAACACCATCTTGTCTCAATTCCTCAATCCCATCTCT 365  
Qy 135 ValGlnAsnLeuGluTyrLeuTyrMetGlyArgGluAsnThrMetProAspLeuAspMet 154  
Db 366 GTTCCCTTCTGGAGATTGACCTTACAGTGATAGAGACTATGTTCTTAACCTTAACCT 425  
Qy 155 LeuAsnGluSerProArgLeuPheAlaGlyHisIleProTyrSerLeuLeuProAlaSer 174  
Db 426 TTATCTCAT---CCAAGACTCTTCTCCACACATATTCCTTGTGCTTGTGTTGCCAAAATCT 482  
Qy 175 ValLeuLysSerGlyThrLysIlelelelelelelelelelelelelelelelelele 194  
Db 483 GTGAAGAATCACTCCAGAGTTGTGTCTGTGTAGAGATCTTAAGACATCTTTGTT 542  
Qy 195 SerPhe-----TyrLysPheGlyAsnLeuLeuAsnProAspLysLeu-----LeuAsp 210  
Db 543 TCACT-TTACTTTGGCATTTTCAAAACAAGTAAAGACACAGAGTAGAGGACCATCTCCA 601  
Qy 211 LeuGluLysSerValAspIlePheAlaSerGlyIleSerPheCysGlyProGluTyr--- 229  
Db 602 TTGAAGAGGCTTTTGAGAGTTTGTAGAGAGTGTGATCTTATGAGCCCTTTGGAG 661  
Qy 230 AsnPheGlnAlaGluPheThrAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSer 249  
Db 662 CATGTATTAGTATTTGGAGAAAAAATTTGAAAGGCCAGAGAGGTGATGTTATATAAA 721

QY 250 TyrGluGluMetLeuGluLysProValGluAsnValLysLysLeuAlaGluPheMetGly 269  
Db 722 TTTGAAGAAATGAAATGAAACCTAGTTTTATTATTGAAAGAGATTGCTGAGTTCTGGGT 781  
QY 270 CysGlyPheThrAspAspGluGluGlnGlylleValAspGluLysLeuCy 289  
Db 782 TGTCCATCTTTAATGAAGAAATCCAAAGGCGATGATGATATATTAAATTTTGTG 841  
QY 289 sSerPheAspAsn 293  
Db 842 TAGTTTGGAGAC 854

RESULT 15  
LOCUS CG269067  
DEFINITION CGXCR72TV ZM\_0.7\_1.5\_KB Zea mays genomic clone ZMMBMA0650K24,  
genomic survey sequence.  
ACCESSION CG269067  
VERSION CG269067.1 GI:34181208  
KEYWORDS GSS.  
SOURCE Zea mays  
ORGANISM Zea mays  
REFERENCE 1 (bases 1 to 966)  
AUTHORS Whitelaw,C.A., Quackenbush,J., Van Aken,S., Utterback,T.,  
Sternick,A., Frazer,C.M., Budiman,M.A., Bedell,J.A., Rohlfing,T.,  
Citek,R.W., Nurnberg,A., Robbins,D. and Lakey,N.  
TITLE Consortium for Maize Genomics  
JOURNAL Unpublished (2002)  
COMMENT Other GSSs: OGXCR72TH  
Contact: Cathy Whitelaw  
TIGR Medical Center Drive, Rockville, MD 20850, USA  
Tel: 301-838-5843  
Fax: 301-838-0208  
Email: whitelaw@tigr.org  
Seq primer: TF  
Class: sheared ends

FEATURES  
source  
1..966  
/organism="Zea mays"  
/mol\_type="genomic DNA"  
/strain="B73"  
/db\_xref="taxon:4577"  
/clone="ZMMBMA0650K24"  
/clone\_lib="ZM\_0.7\_1.5\_KB"  
/notes="Vector: pBCSK-; Site 1: HincII; 0.7-1.5 kb  
methylation filtered genomic DNA library"

ORIGIN  
Alignment Scores:  
Pred. No.: 2,14e-36 Length: 966  
Score: 426.50 Matches: 108  
Percent Similarity: 53.57% Conservative: 57  
Best Local Similarity: 35.06% Mismatches: 117  
Query Match: 20.71% Indels: 26  
DB: 29 Gaps: 10

US-09-854-122-16 (1-391) x CG269067 (1-966)  
QY 14 LeuIleMetAlaGlylleLeuAlaLeuGluLys-----CysPhe-GlySe 28  
Db 25 TTAATCCACCGGCCATCGTTCTTCTGAACAAAGCCATGGCGCAAGCTGCTCCGGCTC 84  
QY 28 rLysAsnGluGlnGluLys-----GluGluAspSerLysMetTyrIleArgTy 44  
Db 85 CAGCAAGATCGACACCAAGCTCCACCTCAGCGGAGGCGGCTCAACCCAGAGCTA 144  
QY 44 rArgGluIleValSerSerLeuProSerAsnAspTyrTrpGlyAspThrMetArgHeuty 64

Db 145 CGGCGACCTGGTATCCACTCTCCCGGTAGAGCGGGTGGGTATCGCTCGTGCAG---TA 201  
QY 64 rLysGlyPheTrpGlnMetGlyTyrLeuValProGly-----IleMetAlaPh 80  
Db 202 CCAGAACTACTGG-----CTGAACCCCGGAGGCTCCAGAACATCATCCCCGT 249  
QY 80 eGluAspAsnPhelYsAlaArgGluThrAspIleIleLeuThrThrLeuProLysAlaGl 100  
Db 250 GAAAGAGTCTACAAAGCTCGCGCGACGACATCTCTCGCCACCTACCTGAATGCGG 309  
QY 100 yThrThrTrpThrYsAlaLeuThrPheAlaIleLeuThrArgaspValAsnHisPro-- 119  
Db 310 GACCACGTGGTGAAGGCGTGGCCTTCGCATACGACCCCGCGCGCCACCAACGAGC 369  
QY 120 ----SerSerProThrHisProLeuLeuPhePheAsnProHisSerCysValGlnAsnLe 138  
Db 370 CTTCCGCGCGCGCGACCACTCCCTCTCAGGTGCACCCGCGAGGAGGCGGTGCCCACT 429  
QY 138 uGluTrpLeuTrpMetGlyArgGluAsnThrMetProAspLeuAspMetLeuAsnGluSe 158  
Db 430 GGAGGTGCACACGCCAGGGGCGA-----GGCTGGCAGACATCGAGAGGCTG---CCGTC 480  
QY 158 rProArgLeuPheAlaGlyHisIleProTy--SerLeuLeuProAlaSerValLeuLysSe 178  
Db 481 GCCGAGGTGCTCGGAGCGACCTGCCCTCTCCTCTGCTCCCGCGCGGTGGCCACCTC 540  
QY 178 rGlyThrLysIlelleAsnIleSerArgAsnArgLysSerThrPheValSerPheTrpLy 198  
Db 541 CGGCTCGCGCGTCTGCTGTACTCTGCGCGACGCCCAAGGACGTGTCTGCTCGCTCGCA 600  
QY 198 sPheGlyAsnLeuIleAsnProAspLysLeu---LeuAspLeuGluLysSerValAspIl 217  
Db 601 CTTCTGTAAGGCGATGCGCGGGGCGAGTCCCGGTGGAGTGGACGCCGCGGTGGCCAT 660  
QY 217 ePheAlaSerGlylleSerPheCysGlyProGluTrpAsnPheGlnAlaGluPhe---Th 236  
Db 661 GTTCTGCGAGGGGGTCTCGCCTTTTCGGGCGCGTCTGGAGCACTACTCTGAGTACTGGA 720  
QY 236 rAsnAlaAlaSerThrAsnSerAsnLeuLeuLeuSerTyrGluGluMetLeuGluLy 256  
Db 721 GGAGAGCTGGCGAGGCGCTGAGCGGGTCTTTCTCTAGGTACGAGGAGATGGTGGCGGA 780  
QY 256 sProValGluAsnValLysLysLeuAlaGluPheMetGlyCysGlyPheThrAspAspGl 276  
Db 781 CCGGTCGCTGCGGTGAGGAGCGTCCGCGGCTTCTTCGCCGCTCCGCTTACCGACGAGA 840  
QY 276 uGluLysGlnGlylleValAspGluIleValLysLeuCysSerPheAspAsnLeuLysAs 296  
Db 841 AGAGGGCGAGGGGTCCCGGAGGAGATTGTGAGGCTGTGCAGCTTCGAGATGCTGAGCGG 900  
QY 296 nGlnGlnValAsnLysAsnGly 303  
Db 901 CTTGGAAAAGCATCGGACCGGA 922

Search completed: March 20, 2004, 06:21:51  
Job time : 3322 secs